

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

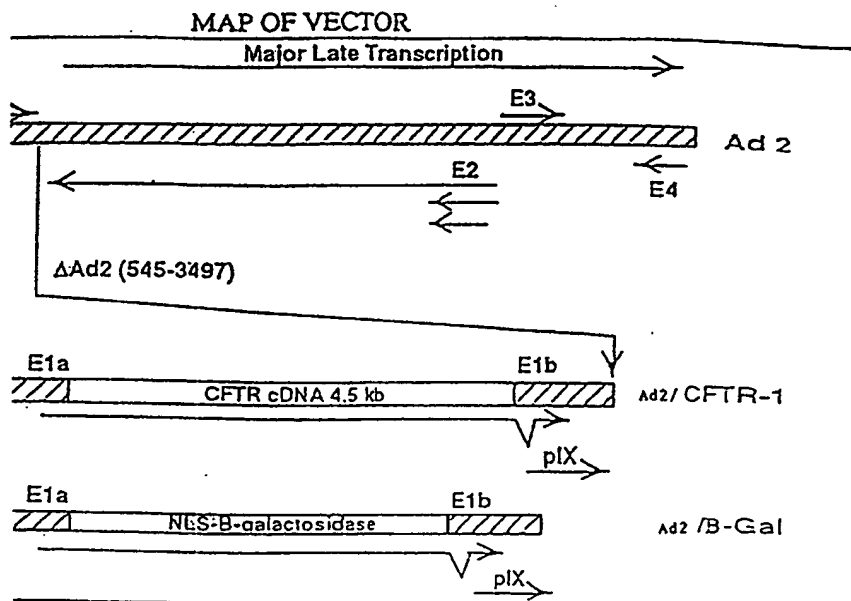
(51) International Patent Classification <sup>5</sup> : C12N 15/86, 15/12, A61K 48/00		A2	(11) International Publication Number: WO 94/12649
			(43) International Publication Date: 9 June 1994 (09.06.94)
(21) International Application Number: PCT/US93/11667			(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  Published Without international search report and to be republished upon receipt of that report.
(22) International Filing Date: 2 December 1993 (02.12.93)			
(30) Priority Data: 07/985,478      3 December 1992 (03.12.92)      US 08/130,682      1 October 1993 (01.10.93)      US 08/136,742      13 October 1993 (13.10.93)      US			
(71) Applicant: GENZYME CORPORATION [US/US]; One Kendall Square, Cambridge, MA 02139 (US).			
(72) Inventors: GREGORY, Richard, J.; 4789 Gateshead Road, Carlsbad, CA 92008 (US). ARMENTANO, Donna; 33 Carver Road, Watertown, MA 02172 (US). COUTURE, Larry, A.; 67 Circle Drive, Framingham, MA 01701 (US). SMITH, Alan, E.; 88 Cleveland Road, Wellesley, MA 02181 (US).			
(74) Agents: HANLEY, Elizabeth, A. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).			

(54) Title: GENE THERAPY FOR CYSTIC FIBROSIS

## (57) Abstract

Gene Therapy vectors, which are especially useful for cystic fibrosis, and methods for using the vectors are disclosed. In preferred embodiments, the vectors are adenovirus-based. Advantages of adenovirus-based vectors for gene therapy are that they appear to be relatively safe and can be manipulated to encode the desired gene product and at the same time are inactivated in terms of their ability to replicate in a normal lytic viral life cycle. Additionally, adenovirus has a natural tropism for airway epithelia. Therefore, adenovirus-based vectors are particularly preferred for respiratory gene therapy applications such as gene therapy for cystic fibrosis. In one embodiment, the adenovirus-based gene therapy vector comprises an adenovirus 2 serotype genome in which the E1a and E1b regions of the genome, which are involved in

early stages of viral replication have been deleted and replaced by genetic material of interest (e.g., DNA encoding the cystic fibrosis transmembrane regulator protein). In another embodiment, the adenovirus-based therapy vector is a pseudo-adenovirus (PAV). PAVs contain no potentially harmful viral genes, have a theoretical capacity for foreign material of nearly 36 kb, may be produced in reasonably high titers and maintain the tropism of the parent adenovirus for dividing and non-dividing human target cell types.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## GENE THERAPY FOR CYSTIC FIBROSIS

### Related Applications

This application is a continuation-in-part application of United States Serial Number 08/130,682, filed on October 1, 1993 which is a continuation-in-part application of United States Serial Number 07/985,478, filed on December 2, 1992, which is a continuation-in-part application of United States Serial Number 07/613,592, filed on November 15, 1990, which is in turn a continuation-in-part application of United States Serial Number 07/589,295, filed on September 27, 1990, which is itself a continuation-in-part application of United States Serial Number 07/488,307, filed on March 5, 1990. The contents of all of the above co-pending patent applications are incorporated herein by reference. Definitions of language or terms not provided in the present application are the same as those set forth in the copending applications. Any reagents or materials used in the examples of the present application whose source is not expressly identified also is the same as those described in the copending application, e.g.,  $\Delta F508$  CFTR gene and CFTR antibodies.

### Background of the Invention

Cystic Fibrosis (CF) is the most common fatal genetic disease in humans (Boat, T.F. et al. in *The Metabolic Basis of Inherited Diseases* (Scriver, C.R. et al. eds., McGraw-Hill, New York (1989)). Approximately one in every 2,500 infants in the United States is born with the disease. At the present time, there are approximately 30,000 CF patients in the United States. Despite current standard therapy, the median age of survival is only 26 years. Disease of the pulmonary airways is the major cause of morbidity and is responsible for 95% of the mortality. The first manifestation of lung disease is often a cough, followed by progressive dyspnea. Tenacious sputum becomes purulent because of colonization of *Staphylococcus* and then with *Pseudomonas*. Chronic bronchitis and bronchiectasis can be partially treated with current therapy, but the course is punctuated by increasingly frequent exacerbations of the pulmonary disease. As the disease progresses, the patient's activity is progressively limited. End-stage lung disease is heralded by increasing hypoxemia, pulmonary hypertension, and cor pulmonale.

The upper airways of the nose and sinuses are also involved in CF. Most patients with CF develop chronic sinusitis. Nasal polyps occur in 15-20% of patients and are common by the second decade of life. Gastrointestinal problems are also frequent in CF; infants may suffer meconium ileus. Exocrine pancreatic insufficiency, which produces symptoms of malabsorption, is present in the large majority of patients with CF. Males are almost uniformly infertile and fertility is decreased in females.

Based on both genetic and molecular analyses, a gene associated with CF was isolated as part of 21 individual cDNA clones and its protein product predicted (Kerem, B.S. et al. (1989) *Science* 245:1073-1080; Riordan, J.R. et al. (1989) *Science* 245:1066-1073;

Rommens, J.M. et al. (1989) *Science* 245:1059-1065)). United States Serial Number 07/488,307 describes the construction of the gene into a continuous strand, expression of the gene as a functional protein and confirmation that mutations of the gene are responsible for CF. (See also Gregory, R.J. et al. (1990) *Nature* 347:382-386; Rich, D.P. et al. (1990) *Nature* 347:358-362). The co-pending patent application also discloses experiments which show that proteins expressed from wild type but not a mutant version of the cDNA complemented the defect in the cAMP regulated chloride channel shown previously to be characteristic of CF.

The protein product of the CF associated gene is called the cystic fibrosis transmembrane conductance regulator (CFTR) (Riordan, J.R. et al. (1989) *Science* 245:1066-1073). CFTR is a protein of approximately 1480 amino acids made up of two repeated elements, each comprising six transmembrane segments and a nucleotide binding domain. The two repeats are separated by a large, polar, so-called R-domain containing multiple potential phosphorylation sites. Based on its predicted domain structure, CFTR is a member of a class of related proteins which includes the multi-drug resistance (MDR) or P-glycoprotein, bovine adenylyl cyclase, the yeast STE6 protein as well as several bacterial amino acid transport proteins (Riordan, J.R. et al. (1989) *Science* 245:1066-1073; Hyde, S.C. et al. (1990) *Nature* 346:362-365). Proteins in this group, characteristically, are involved in pumping molecules into or out of cells.

CFTR has been postulated to regulate the outward flow of anions from epithelial cells in response to phosphorylation by cyclic AMP-dependent protein kinase or protein kinase C (Riordan, J.R. et al. (1989) *Science* 245:1066-1073; Welsh, 1986; Frizzell, R.A. et al. (1986) *Science* 233:558-560; Welsh, M.J. and Liedtke, C.M. (1986) *Nature* 322:467; Li, M. et al. (1988) *Nature* 331:358-360; Huang, T-C. et al. (1989) *Science* 244:1351-1353).

Sequence analysis of the CFTR gene of CF chromosomes has revealed a variety of mutations (Cutting, G.R. et al. (1990) *Nature* 346:366-369; Dean, M. et al. (1990) *Cell* 61:863-870; and Kerem, B-S. et al. (1989) *Science* 245:1073-1080; Kerem, B-S. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:8447-8451). Population studies have indicated that the most common CF mutation, a deletion of the 3 nucleotides that encode phenylalanine at position 508 of the CFTR amino acid sequence ( $\Delta F508$ ), is associated with approximately 70% of the cases of cystic fibrosis. This mutation results in the failure of an epithelial cell chloride channel to respond to cAMP (Frizzell R.A. et al. (1986) *Science* 233:558-560; Welsh, M.J. (1986) *Science* 232:1648-1650.; Li, M. et al. (1988) *Nature* 331:358-360; Quinton, P.M. (1989) *Clin. Chem.* 35:726-730). In airway cells, this leads to an imbalance in ion and fluid transport. It is widely believed that this causes abnormal mucus secretion, and ultimately results in pulmonary infection and epithelial cell damage.

Studies on the biosynthesis (Cheng, S.H. et al. (1990) *Cell* 63:827-834; Gregory, R.J. et al. (1991) *Mol. Cell Biol.* 11:3886-3893) and localization (Denning, G.M. et al. (1992) *J. Cell Biol.* 118:551-559) of CFTR  $\Delta F508$ , as well as other CFTR mutants, indicate that many CFTR mutant proteins are not processed correctly and, as a result, are not delivered to the

plasma membrane (Gregory, R.J. et al. (1991) *Mol. Cell Biol.* 11:3886-3893). These conclusions are consistent with earlier functional studies which failed to detect cAMP-stimulated Cl<sup>-</sup> channels in cells expressing CFTR  $\Delta$ F508 (Rich, D.P. et al. (1990) *Nature* 347:358-363; Anderson, M.P. et al. (1991) *Science* 251:679-682).

- 5 To date, the primary objectives of treatment for CF have been to control infection, promote mucus clearance, and improve nutrition (Boat, T.F. et al. in *The Metabolic Basis of Inherited Diseases* (Scriver, C.R. et al. eds., McGraw-Hill, New York (1989)). Intensive antibiotic use and a program of postural drainage with chest percussion are the mainstays of therapy. However, as the disease progresses, frequent hospitalizations are required.
- 10 Nutritional regimens include pancreatic enzymes and fat-soluble vitamins. Bronchodilators are used at times. Corticosteroids have been used to reduce inflammation, but they may produce significant adverse effects and their benefits are not certain. In extreme cases, lung transplantation is sometimes attempted (Marshall, S. et al. (1990) *Chest* 98:1488).

- Most efforts to develop new therapies for CF have focused on the pulmonary
- 15 complications. Because CF mucus consists of a high concentration of DNA, derived from lysed neutrophils, one approach has been to develop recombinant human DNase (Shak, S. et al. (1990) *Proc. Natl. Sci. Acad USA* 87:9188). Preliminary reports suggest that aerosolized enzyme may be effective in reducing the viscosity of mucus. This could be helpful in clearing the airways of obstruction and perhaps in reducing infections. In an attempt to limit
- 20 damage caused by an excess of neutrophil derived elastase, protease inhibitors have been tested. For example, alpha-1-antitrypsin purified from human plasma has been aerosolized to deliver enzyme activity to lungs of CF patients (McElvaney, N. et al. (1991) *The Lancet* 337:392). Another approach would be the use of agents to inhibit the action of oxidants derived from neutrophils. Although biochemical parameters have been successfully
- 25 measured, the long term beneficial effects of these treatments have not been established.

- Using a different rationale, other investigators have attempted to use pharmacological agents to reverse the abnormally decreased chloride secretion and increased sodium absorption in CF airways. Defective electrolyte transport by airway epithelia is thought to alter the composition of the respiratory secretions and mucus (Boat, T.F. et al. in *The*
- 30 *Metabolic Basis of Inherited Diseases* (Scriver, C.R. et al. eds., McGraw-Hill, New York (1989); Quinton, P.M. (1990) *FASEB J.* 4:2709-2717). Hence, pharmacological treatments aimed at correcting the abnormalities in electrolyte transport could be beneficial. Trials are in progress with aerosolized versions of the drug amiloride; amiloride is a diuretic that inhibits sodium channels, thereby inhibiting sodium absorption. Initial results indicate that the drug
- 35 is safe and suggest a slight change in the rate of disease progression, as measured by lung function tests (Knowles, M. et al. (1990) *N. Eng. J. Med.* 322: 1189-1194; App, E. (1990) *Am. Rev. Respir. Dis.* 141:605). Nucleotides, such as ATP or UTP, stimulate purinergic receptors in the airway epithelium. As a result, they open a class of chloride channel that is different from CFTR chloride channels. *In vitro* studies indicate that ATP and UTP can stimulate

chloride secretion (Knowles, M. et al. (1991) *N. Eng. J. Med.* 325:533). Preliminary trials to test the ability of nucleotides to stimulate secretion *in vivo*, and thereby correct the electrolyte transport abnormalities are underway.

5 Despite progress in therapy, cystic fibrosis remains a lethal disease, and no current therapy treats the basic defect. However, two general approaches may prove feasible. These are: 1) protein replacement therapy to deliver the wild type protein to patients to augment their defective protein, and; 2) gene replacement therapy to deliver wild type copies of the CF associated gene. Since the most life threatening manifestations of CF involve pulmonary complications, epithelial cells of the upper airways are appropriate target cells for therapy.

10 The feasibility of gene therapy has been established by introducing a wild type cDNA into epithelial cells from a CF patient and demonstrating complementation of the hallmark defect in chloride ion transport (Rich, D.P. et al. (1990) *Nature* 347:358-363 ). This initial work involved cells in tissue culture, however, subsequent work has shown that to deliver the gene to the airways of whole animals, defective adenoviruses may be useful (Rosenfeld, 15 (1992) *Cell* 68:143-155). However, the safety and effectiveness of using defective adenoviruses remain to be demonstrated.

#### Summary of the Invention

In general, the instant invention relates to vectors for transferring selected genetic material of interest (e.g., DNA or RNA) to cells *in vivo*. In preferred embodiments, the 20 vectors are adenovirus-based. Advantages of adenovirus-based vectors for gene therapy are that they appear to be relatively safe and can be manipulated to encode the desired gene product and at the same time are inactivated in terms of their ability to replicate in a normal lytic viral life cycle. Additionally, adenovirus has a natural tropism for airway epithelia. 25 Therefore, adenovirus-based vectors are particularly preferred for respiratory gene therapy applications such as gene therapy for cystic fibrosis.

In one embodiment, the adenovirus-based gene therapy vector comprises an adenovirus 2 serotype genome in which the Ela and Elb regions of the genome, which are involved in early stages of viral replication have been deleted and replaced by genetic 30 material of interest (e.g., DNA encoding the cystic fibrosis transmembrane regulator protein).

In another embodiment, the adenovirus-based therapy vector is a pseudo-adenovirus (PAV). PAVs contain no potentially harmful viral genes, have a theoretical capacity for foreign material of nearly 36 kb, may be produced in reasonably high titers and maintain the tropism of the parent adenovirus for dividing and non-dividing human target cell types. 35 PAVs comprise adenovirus inverted terminal repeats and the minimal sequences of a wild-type adenovirus type 2 genome necessary for efficient replication and packaging by a helper virus and genetic material of interest. In a preferred embodiment, the PAV contains adenovirus 2 sequences.

In a further embodiment, the adenovirus-based gene therapy vector contains the open reading frame 6 (ORF6) of adenoviral early region 4 (E4) from the E4 promoter and is deleted for all other E4 open reading frames. Optionally, this vector can include deletions in the E1 and/or E3 regions. Alternatively, the adenovirus-based gene therapy vector contains the open reading frame 3 (ORF3) of adenoviral E4 from the E4 promoter and is deleted for all other E4 open reading frames. Again, optionally, this vector can include deletions in the E1 and/or E3 regions. The deletion of non-essential open reading frames of E4 increases the cloning capacity by approximately 2 kb without significantly reducing the viability of the virus in cell culture. In combination with deletions in the E1 and/or E3 regions of adenovirus vectors, the theoretical insert capacity of the resultant vectors is increased to 8-9 kb.

The invention also relates to methods of gene therapy using the disclosed vectors and genetically engineered cells produced by the method.

#### **Brief Description of the Tables and Drawings**

Further understanding of the invention may be had by reference to the tables and figures wherein:

Table I shows CFTR mutants wherein the known association with CF (Y, yes or N, no), exon localization, domain location and presence (+) or absence (-) of bands A, B, and C of mutant CFTR species is shown. TM6, indicates transmembrane domain 6; NBD nucleotide binding domain; ECD, extracellular domain and Term, termination at 21 codons past residue 1337;

Table II shows the nucleotide sequence of Ad2/CFTR-1;

Table III depicts a nucleotide analysis of Ad2-ORF6/PGK-CFTR;

The convention for naming mutants is first the amino acid normally found at the particular residue, the residue number (Riordan, T.R. et al. (1989) *Science* 245:1066-1073). and the amino acid to which the residue was converted. The single letter amino acid code is used: D, aspartic acid; F, phenylalanine; G, glycine; I, isoleucine; K, lysine; M, methionine; N, asparagine; Q, glutamine; R, arginine; S, serine; W, tryptophan. Thus G551D is a mutant in which glycine 551 is converted to aspartic acid;

Figure 1 shows alignment of CFTR partial cDNA clones used in construction of cDNA containing complete coding sequence of the CFTR, only restriction sites relevant to the DNA constructions described below are shown;

Figure 2 depicts plasmid construction of the CFTR cDNA clone pKK-CFTR1;

Figure 3 depicts plasmid construction of the CFTR cDNA clone pKK-CFTR2;

Figure 4 depicts plasmid construction of the CFTR cDNA clone pSC-CFTR2;

5

Figure 5 shows a plasmid map of the CFTR cDNA clone pSC-CFTR2;

Figure 6 shows the DNA sequence of synthetic DNAs used for insertion of an intron into the CFTR cDNA sequence, with the relevant restriction endonuclease sites and  
10 nucleotide positions noted;

Figures 7A and 7B depict plasmid construction of the CFTR cDNA clone pKK-CFTR3;

Figure 8 shows a plasmid map of the CFTR cDNA pKK-CFTR3 containing an intron between nucleotides 1716 and 1717;

15

Figure 9 shows treatment of CFTR with glycosidases;

Figures 10A and 10B show an analysis of CFTR expressed from COS-7 transfected cells;

20

Figures 11A and 11B show pulse-chase labeling of wild type and  $\Delta F508$  mutant CFTR in COS-7 transfected cells;

25

Figures 12A-12D show immunolocalization of wild type and  $\Delta F508$  mutant CFTR; and COS-7 cells transfected with pMT-CFTR or pMT-CFTR- $\Delta F508$ ;

Figure 13 shows an analysis of mutant forms of CFTR;

30

Figure 14 shows a map of the first generation adenovirus based vector encoding CFTR (Ad2/CFTR-1);

Figure 15 shows the plasmid construction of the Ad2/CFTR-1 vector;

35

Figure 16 shows an example of UV fluorescence from an agarose gel electrophoresis of products of nested RT-PCR from lung homogenates of cotton rats which received Ad2/CFTR-1. The gel demonstrates that the homogenates were positive for virally-encoded CFTR mRNA;



Figure 17 shows an example of UV fluorescence from an agarose gel electrophoresis of products of nested RT-PCR from organ homogenates of cotton rats. The gel demonstrates that all organs of the infected rats were negative for Ad2/CFTR with the exception of the small bowel;

Figures 18A and 18B show differential cell analyses of bronchoalveolar lavage specimens from control and infected rats. These data demonstrate that none of the rats treated with Ad2/CFTR-1 had a change in the total or differential white blood cell count 4, 10, and 14 days after infection (Figure 18A) and 3, 7, and 14 days after infection (Figure 18B);

Figure 19 shows hematoxylin and eosin stained sections of cotton rat tracheas from both treated and control rats sacrificed at different time points after infection with Ad2/CFTR-1. The sections demonstrate that there were no observable differences between the treated and control rats;

Figures 20A and 20B show examples of UV fluorescence from an agarose gel electrophoresis, stained with ethidium bromide, of products of RT-PCR from nasal brushings of Rhesus monkeys after application of Ad2/CFTR-1 or Ad2/ $\beta$ -Gal;

Figure 21 shows lights microscopy and immunocytochemistry from monkey nasal brushings. The microscopy revealed that there was a positive reaction when nasal epithelial cells from monkeys exposed to Ad2/CFTR-1 were stained with antibodies to CFTR;

Figure 22 shows immunocytochemistry of monkey nasal turbinate biopsies. This microscopy reveals increased immunofluorescence at the apical membrane of the surface epithelium from biopsies obtained from monkeys treated with Ad2/CFTR-1 over that seen at the apical membrane of the surface epithelium from biopsies obtained from control monkeys;

Figures 23A-23D show serum antibody titers in Rhesus monkeys after three vector administrations. These graphs demonstrate that all three monkeys treated with Ad2/CFTR-1 developed antibodies against adenovirus;

Figure 24 shows hematoxylin and eosin stained sections from monkey medial turbinate biopsies. These sections demonstrate that turbinate biopsy specimens from control monkeys could not be differentiated from those from monkeys treated with Ad2/CFTR-1 when reviewed by an independent pathologist;

Figures 25A-25I show photomicrographs of human nasal mucosa immediately before, during, and after Ad2/CFTR-1 application. These photomicrographs demonstrate that inspection of the nasal mucosa showed mild to moderate erythema, edema, and exudate in patients treated with Ad2/CFTR-1 (Figures 25A-25C) and in control patients (Figures 25G-25I). These changes were probably due to local anesthesia and vasoconstriction because when an additional patient was exposed to Ad2/CFTR in a method which did not require the use of local anesthesia or vasoconstriction, there were no symptoms and the nasal mucosa appeared normal (Figures 25D-25F);

Figure 26 shows a photomicrograph of a hematoxylin and eosin stained biopsy of human nasal mucosa obtained from the third patient three days after Ad2/CFTR-1 administration. This section shows a morphology consistent with CF, i.e., a thickened basement membrane and occasional morphonuclear cells in the submucosa, but no abnormalities that could be attributed to the adenovirus vector;

Figure 27 shows transepithelial voltage ( $V_t$ ) across the nasal epithelium of a normal human subject. Amiloride ( $\mu\text{M}$ ) and terbutaline ( $\mu\text{M}$ ) were perfused onto the mucosal surface beginning at the times indicated. Under basal conditions ( $V_t$ ) was electrically negative. Perfusion of amiloride onto the mucosal surface inhibited ( $V_t$ ) by blocking apical  $\text{Na}^+$  channels;

Figures 28A and 28B show transepithelial voltage ( $V_t$ ) across the nasal epithelium of normal human subjects (Figure 28A) and patients with CF (Figure 28B). Values were obtained under basal conditions, during perfusion with amiloride ( $\mu\text{M}$ ), and during perfusion of amiloride plus terbutaline ( $\mu\text{M}$ ) onto the mucosal surface. Data are from seven normal subjects and nine patients with CF. In patients with CF, ( $V_t$ ) was more electrically negative than in normal subjects (Figure 28B). Amiloride inhibited ( $V_t$ ) in CF patients, as it did in normal subjects. However,  $V_t$  failed to hyperpolarize when terbutaline was perfused onto the epithelium in the presence of amiloride. Instead, ( $V_t$ ) either did not change or became less negative, a result very different from that observed in normal subjects;

Figures 29A and 29B show transepithelial voltage ( $V_t$ ) across the nasal epithelium of a third patient before (Figure 29A) and after (Figure 29B) administration of approximately 25 MOI of Ad2/CFTR-1. Amiloride and terbutaline were perfused onto the mucosal surface beginning at the times indicated. Figure 29A shows an example from the third patient before treatment. Figure 29B shows that in contrast to the response before Ad2/CFTR-1 was applied, after virus replication, in the presence of amiloride, terbutaline stimulated  $V_t$ ;

Figures 30A-30F show the time of course changes in transepithelial electrical properties before and after administration of Ad2/CFTR-1. Figures 30A and 30B are from the first patient who received approximately 1 MOI; Figures 30C and 30D are from the second patient who received approximately 3 MOI; and Figures 30E and 30F are from the third patient who received approximately 25 MOI. Figures 30A, 30C, and 30E show values of basal transepithelial voltage ( $V_t$ ) and Figures 30B, 30D, and 30F show the change in transepithelial voltage ( $\Delta V_t$ ) following perfusion of terbutaline in the presence of amiloride. Day zero indicates the day of Ad2/CFTR-1 administration. Figures 30A, 30C, and 30E show the time course of changes in basal  $V_t$  for all three patients. The decrease in basal  $V_t$  suggests that application of Ad2/CFTR-1 corrected the CF electrolyte transport defect in nasal epithelium of all three patients. Additional evidence came from an examination of the response to terbutaline. Figures 30B, 30D, and 30F show the time course of the response. These data indicate that Ad2/CFTR-1 corrected the CF defect in  $Cl^-$  transport;

Figure 31 shows the time course of changes in transepithelial electrical properties before and after administration of saline instead of Ad2/CFTR-1 to CF patients. Day zero indicates the time of mock administration. The top graph shows basal transepithelial voltage ( $V_t$ ) and the bottom graph shows the change in transepithelial voltage following perfusion with terbutaline in the presence of amiloride ( $\Delta V_t$ ). Closed symbols are data from two patients that received local anesthetic/vasoconstriction and placement of the applicator for thirty minutes. Open symbol is data from a patient that received local anesthetic/vasoconstriction, but not placement of the applicator. Symptomatic changes and physical findings were the same as those observed in CF patients treated with a similar administration procedure and Ad2/CFTR-1;

Figure 32 shows a map of the second generation adenovirus based vector, PAV;

Figure 33 shows the plasmid construction of a second generation adenoviral vector 6 (Ad E4 ORF6);

Figure 34 is a schematic of Ad2-ORF6/PGK-CFTR which differs from Ad2/CFTR in that the latter utilized the endogenous Ela promoter, had no poly A addition signal directly downstream of CFTR and retained an intact E4 region;

Figure 35 shows short-circuit currents from human CF nasal polyp epithelial cells infected with Ad2-ORF6/PGK-CFTR at multiplicities of 0.3, 3, and 50. At the indicated times: (1) 10  $\mu M$  amiloride, (2) cAMP agonists (10  $\mu M$  forskolin and 100  $\mu M$  IBMX, and (3) 1 mM diphenylamine-2-carboxylate were added to the mucosal solution;

Figures 36A-36D show immunocytochemistry of nasal brushings by laser scanning microscopy of the Rhesus monkey C, before infection (36A) and on 7 days (36B); 24 (36C); and 38 (36D) after the first infection with Ad2-ORF6/PGK-CFTR;

Figures 37A-37D show immunocytochemistry of nasal brushings by laser scanning microscopy of Rhesus monkey D, before infection (37A) and on days 7 (37B); 24 (37C); and 48 (37D) after the first infection with Ad2-ORF6/PGK-CFTR;

Figures 38A-38D show immunocytochemistry of nasal brushings by laser scanning microscopy of the Rhesus monkey E, before infection (38A) and on days 7 (38B); 24 (38C); and 48 (38D) after the first infection with Ad2-ORF6/PGK-CFTR;

Figures 39A-39C show summaries of the clinical signs (or lack thereof) of infection with Ad2-ORF6/PGK-CFTR;

Figures 40A-40C shows a summary of blood counts, sedimentation rate, and clinical chemistries after infection with Ad2-ORF6/PGK-CFTR for monkeys C, D, and E. There was no evidence of a systemic inflammatory response or other abnormalities of the clinical chemistries;

Figure 41 shows summaries of white blood cells counts in monkeys C, D, and E after infection with Ad2-ORF6/PGK-CFTR. These data indicate that the administration of Ad2-ORF6/PGK-CFTR caused no change in the distribution and number of inflammatory cells at any of the time points following viral administration;

Figure 42 shows histology of submucosal biopsy performed on Rhesus monkey C on day 4 after the second viral instillation of Ad2-ORF6/PGK-CFTR. Hematoxylin and eosin stain revealed no evidence of inflammation or cytopathic changes;

Figure 43 shows histology of submucosal biopsy performed on Rhesus monkey D on day 11 after the second viral instillation of Ad2-ORF6/PGK-CFTR. Hematoxylin and eosin stain revealed no evidence of inflammation or cytopathic changes;

Figure 44 shows histology of submucosal biopsy performed on Rhesus monkey E on day 18 after the second viral instillation of Ad2-ORF6/PGK-CFTR. Hematoxylin and eosin stain revealed no evidence of inflammation or cytopathic changes; and

- 10.1 -

Figures 45A-45C show antibody titers to adenovirus prior to and after the first and second administrations of Ad2-ORF6/PGK-CFTR. Prior to administration of Ad2-ORF6/PGK-

CFTR, the monkeys had received instillations of Ad2/CFTR-1. Antibody titers measured by ELISA rose within one week after the first and second administrations of Ad2-ORF6/PGK-CFTR. Serum neutralizing antibodies also rose within a week after viral administration and peaked at day 24. No anti-adenoviral antibodies were detected by ELISA or neutralizing assay in nasal washings of any of the monkeys.

### **Detailed Description and Best Mode**

#### **Gene Therapy**

As used herein, the phrase "gene therapy" refers to the transfer of genetic material (e.g., DNA or RNA) of interest into a host to treat or prevent a genetic or acquired disease or condition. The genetic material of interest encodes a product (e.g., a protein polypeptide, peptide or functional RNA) whose production *in vivo* is desired. For example, the genetic material of interest can encode a hormone, receptor, enzyme or (poly) peptide of therapeutic value. Examples of genetic material of interest include DNA encoding: the cystic fibrosis transmembrane regulator (CFTR), Factor VIII, low density lipoprotein receptor, beta-galactosidase, alpha-galactosidase, beta-glucocerebrosidase, insulin, parathyroid hormone, and alpha-1-antitrypsin.

Although the potential for gene therapy to treat genetic diseases has been appreciated for many years, it is only recently that such approaches have become practical with the treatment of two patients with adenosine deaminase deficiency. The protocol consists of removing lymphocytes from the patients, stimulating them to grow in tissue culture, infecting them with an appropriately engineered retrovirus followed by reintroduction of the cells into the patient (Kantoff, P. et al. (1987) *J. Exp. Med.* 166:219). Initial results of treatment are very encouraging. With the approval of a number of other human gene therapy protocols for limited clinical use, and with the demonstration of the feasibility of complementing the CF defect by gene transfer, gene therapy for CF appears a very viable option.

The concept of gene replacement therapy for cystic fibrosis is very simple; a preparation of CFTR coding sequences in some suitable vector in a viral or other carrier delivered directly to the airways of CF patients. Since disease of the pulmonary airways is the major cause of morbidity and is responsible for 95% of mortality, airway epithelial cells are preferred target cells for CF gene therapy. The first generation of CF gene therapy is likely to be transient and to require repeated delivery to the airways. Eventually, however, gene therapy may offer a cure for CF when the identity of the precursor or stem cell to air epithelial cells becomes known. If DNA were incorporated into airway stem cells, all subsequent generations of such cells would make authentic CFTR from the integrated sequences and would correct the physiological defect almost irrespective of the biochemical basis of the action of CFTR.

Although simple in concept, scientific and clinical problems face approaches to gene therapy, not least of these being that CF requires an *in vivo* approach while all gene therapy treatments in humans to date have involved *ex vivo* treatment of cells taken from the patient followed by reintroduction.

5 One major obstacle to be overcome before gene therapy becomes a viable treatment approach for CF is the development of appropriate vectors to infect tissue manifesting the disease and deliver the therapeutic CFTR gene. Since viruses have evolved very efficient means to introduce their nucleic acid into cells, many approaches to gene therapy make use of engineered defective viruses. However, the use of viruses *in vivo* raises safety concerns.

10 Although potentially safer, the use of simple DNA plasmid constructs containing minimal additional DNA, on the other hand, is often very inefficient and can result in transient protein expression.

The integration of introduced DNA into the host chromosome has advantages in that such DNA will be passed to daughter cells. In some circumstances, integrated DNA may  
15 also lead to high or more sustained expression. However, integration often, perhaps always, requires cellular DNA replication in order to occur. This is certainly the case with the present generation of retroviruses. This limits the use of such viruses to circumstances where cell division occurs in a high proportion of cells. For cells cultured *in vitro*, this is seldom a problem, however, the cells of the airway are reported to divide only infrequently  
20 (Kawanami, O. et al. (1979) *An. Rev. Respir. Dis.* 120:595). The use of retroviruses in CF will probably require damaging the airways (by agents such as SO<sub>2</sub> or O<sub>3</sub>) to induce cell division. This may prove impracticable in CF patients.

Even if efficient DNA integration could be achieved using viruses, the human genome contains elements involved in the regulation of cellular growth only a small fraction of which  
25 are presently identified. By integrating adjacent to an element such as a proto-oncogene or an anti-oncogene, activation or inactivation of that element could occur leading to uncontrolled growth of the altered cell. It is considered likely that several such activation/inactivation steps are usually required in any one cell to induce uncontrolled proliferation (R.A. Weinberg (1989) *Cancer Research* 49:3713 ), which may reduce somewhat the potential risk. On the  
30 other hand, insertional mutagenesis leading to tumor formation is certainly known in animals with some nondefective retroviruses (R.A. Weinberg, *supra*; Payne, G.S. et al. (1982) *Nature* 295:209), and the large numbers of potential integrations occurring during the lifetime of a patient treated repeatedly *in vivo* with retroviruses must raise concerns on the safety of such a procedure.

35 In addition to the potential problems associated with viral DNA integration, a number of additional safety issues arise. Many patients may have preexisting antibodies to some of the viruses that are candidates for vectors, for example, adenoviruses. In addition, repeated use of such vectors might induce an immune response. The use of defective viral vectors

may alleviate this problem somewhat, because the vectors will not lead to productive viral life cycles generating infected cells, cell lysis or large numbers of progeny viruses.

Other issues associated with the use of viruses are the possibility of recombination with related viruses naturally infecting the treated patient, complementation of the viral defects by simultaneous expression of wild type virus proteins and containment of aerosols of the engineered viruses.

Gene therapy approaches to CF will face many of the same clinical challenges at protein therapy. These include the inaccessibility of airway epithelium caused by mucus build-up and the hostile nature of the environment in CF airways which may inactivate viruses/vectors. Elements of the vector carriers may be immunogenic and introduction of the DNA may be inefficient. These problems, as with protein therapy, are exacerbated by the absence of a good animal model for the disease nor a simple clinical end point to measure the efficacy of treatment.

#### 15 CF Gene Therapy Vectors - Possible Options

Retroviruses - Although defective retroviruses are the best characterized system and so far the only one approved for use in human gene therapy (Miller, A.D. (1990) *Blood* 76:271), the major issue in relation to CF is the requirement for dividing cells to achieve DNA integration and gene expression. Were conditions found to induce airway cell division, the *in vivo* application of retroviruses, especially if repeated over many years, would necessitate assessment of the safety aspects of insertional mutagenesis in this context.

Adeno-Associated Virus - (AAV) is a naturally occurring defective virus that requires other viruses such as adenoviruses or herpes viruses as helper viruses (Muzyczka, N. (1992) in *Current Topics in Microbiology and Immunology* 158:97). It is also one of the few viruses that may integrate its DNA into non-dividing cells, although this is not yet certain. Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate, but space for exogenous DNA is limited to about 4.5 kb. CFTR DNA may be towards the upper limit of packaging. Furthermore, the packaging process itself is presently inefficient and safety issues such as immunogenicity, complementation and containment will also apply to AAV. Nevertheless, this system is sufficiently promising to warrant further study.

Plasmid DNA - Naked plasmid can be introduced into muscle cells by injection into the tissue. Expression can extend over many months but the number of positive cells is low (Wolff, J. et al. (1989) *Science* 247:1465). Cationic lipids aid introduction of DNA into some cells in culture (Felgner, P. and Ringold, G.M. (1989) *Nature* 337:387). Injection of cationic lipid plasmid DNA complexes into the circulation of mice has been shown to result in expression of the DNA in lung (Brigham, K. et al. (1989) *Am. J. Med. Sci.* 298:278).



Instillation of cationic lipid plasmid DNA into lung also leads to expression in epithelial cells but the efficiency of expression is relatively low and transient (Hazinski, T.A. et al. (1991) *Am. J. Respir., Cell Mol. Biol.* 4:206). One advantage of the use of plasmid DNA is that it can be introduced into non-replicating cells. However, the use of plasmid DNA in the CF airway environment, which already contains high concentrations of endogenous DNA may be problematic.

Receptor Mediated Entry - In an effort to improve the efficiency of plasmid DNA uptake, attempts have been made to utilize receptor-mediated endocytosis as an entry mechanisms and to protect DNA in complexes with polylysine (Wu, G. and Wu, C.H. (1988) *J. Biol. Chem.* 263:14621). One potential problem with this approach is that the incoming plasmid DNA enters the pathway leading from endosome to lysosome, where much incoming material is degraded. One solution to this problem is the use of transferrin DNA-polylysine complexes linked to adenovirus capsids (Curiel, D.T. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8850). The latter enter efficiently but have the added advantage of naturally disrupting the endosome thereby avoiding shuttling to the lysosome. This approach has promise but at present is relatively transient and suffers from the same potential problems of immunogenicity as other adenovirus based methods.

Adenovirus - Defective adenoviruses at present appear to be a promising approach to CF gene therapy (Berkner, K.L. (1988) *BioTechniques* 6:616). Adenovirus can be manipulated such that it encodes and expresses the desired gene product, (e.g., CFTR), and at the same time is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. In addition, adenovirus has a natural tropism for airway epithelia. The viruses are able to infect quiescent cells as are found in the airways, offering a major advantage over retroviruses. Adenovirus expression is achieved without integration of the viral DNA into the host cell chromosome, thereby alleviating concerns about insertional mutagenesis. Furthermore, adenoviruses have been used as live enteric vaccines for many years with an excellent safety profile (Schwartz, A.R. et al. (1974) *Am. Rev. Respir. Dis.* 109:233-238). Finally, adenovirus mediated gene transfer has been demonstrated in a number of instances including transfer of alpha-1-antitrypsin and CFTR to the lungs of cotton rats (Rosenfeld, M.A. et al. (1991) *Science* 252:431-434; Rosenfeld et al., (1992) *Cell* 68:143-155). Furthermore, extensive studies to attempt to establish adenovirus as a causative agent in human cancer were uniformly negative (Green, M. et al. (1979) *Proc. Natl. Acad. Sci. USA* 76:6606).

The following properties would be desirable in the design of an adenovirus vector to transfer the gene for CFTR to the airway cells of a CF patient. The vector should allow sufficient expression of the CFTR, while producing minimal viral gene expression. There should be minimal viral DNA replication and ideally no virus replication. Finally,

recombination to produce new viral sequences and complementation to allow growth of the defective virus in the patient should be minimized. A first generation adenovirus vector encoding CFTR (Ad2/CFTR), made as described in the following Example 7, achieves most of these goals and was used in the human trials described in Example 10.

5        Figure 14 shows a map of Ad2/CFTR-1. As can be seen from the figure, this first generation virus includes viral DNA derived from the common relatively benign adenovirus 2 serotype. The Ela and Elb regions of the viral genome, which are involved in early stages of viral replication have been deleted. Their removal impairs viral gene expression and viral replication. The protein products of these genes also have immortalizing and transforming  
10       function in some non-permissive cells.

      The CFTR coding sequence is inserted into the viral genome in place of the Ela/Elb region and transcription of the CFTR sequence is driven by the endogenous Ela promoter. This is a moderately strong promoter that is functional in a variety of cells. In contrast to some adenovirus vectors (Rosenfeld, M. et al. (1992) *Cell* 68:143), this adenovirus retains  
15       the E3 viral coding region. As a consequence of the inclusion of E3, the length of the adenovirus-CFTR DNA is greater than that of the wild-type adenovirus. The greater length of the recombinant viral DNA renders it more difficult to package. This means that the growth of the Ad2/CFTR virus is impaired even in permissive cells that provide the missing Ela and Elb functions.

20       The E3 region of the Ad2/CFTR-1 encodes a variety of proteins. One of these proteins, gp19, is believed to interact with and prevent presentation of class I proteins of the major histocompatibility complex (MHC) (Gooding, C.R. and Wold, W.S.M. (1990) *Crit. Rev. Immunol.* 10:53). This property prevents recognition of the infected cells and thus may allow viral latency. The presence of E3 sequences, therefore, has two useful attributes; first,  
25       the large size of the viral DNA renders it doubly defective for replication (i.e., it lacks early functions and is packaged poorly) and second, the absence of MHC presentation could be useful in later applications of Ad2/CFTR-1 in gene therapy involving multiple administrations because it may avoid an immune response to recombinant virus containing cells.

30       Not only are there advantages associated with the presence of E3; there may be disadvantages associated with its absence. Studies of E3 deleted virus in animals have suggested that they result in a more severe pathology (Gingsberg, H.S. et al. (1989) *Proc. Natl. Acad. Sci. (USA)* 86:3823). Furthermore, E3 deleted virus, such as might be obtained by recombination of an E1 plus E3 deleted virus with wild-type virus, is reported to outgrow  
35       wild-type in tissue culture (Barkner, K.L. and Sharp, P. (1983) *Nucleic Acids Research* 11:6003). By contrast, however, a recent report of an E3 replacement vector encoding hepatitis B surface antigen, suggests that when delivered as a live enteric vaccine, such a virus replicates poorly in human compared to wild-type.

The adenovirus vector (Ad2/CFTR-1) and a related virus encoding the marker  $\beta$ -galactosidase (Ad2/ $\beta$ -gal) have been constructed and grown in human 293 cells. These cells contain the E1 region of adenovirus and constitutively express Ela and Elb, which complement the defective adenoviruses by providing the products of the genes deleted from the vector. Because the size of its genome is greater than that of wild-type virus, Ad2/CFTR is relatively difficult to produce.

The Ad2/CFTR-1 virus has been shown to encode CFTR by demonstrating the presence of the protein in 293 cells. The Ad2/ $\beta$ -gal virus was shown to produce its protein in a variety of cell lines grown in tissue culture including a monkey bronchiolar cell line (4MBR-5), primary hamster tracheal epithelial cells, human HeLa, human CF PAC cells (see Example 8) and airway epithelial cells from CF patients (Rich, O. et al. (1990) *Nature* 347:358).

Ad2/CFTR-1 is constructed from adenovirus 2 (Ad2) DNA sequences. Other varieties of adenovirus (e.g., Ad3, Ad5, and Ad7) may also prove useful as gene therapy vectors. This may prove essential if immune response against a single serotype reduces the effectiveness of the therapy.

#### Second Generation Adenoviral Vectors

Adenoviral vectors currently in use retain most ( $\geq 80\%$ ) of the parental viral genetic material leaving their safety untested and in doubt. Second-generation vector systems containing minimal adenoviral regulatory, packaging and replication sequences have therefore been developed.

Pseudo-Adenovirus Vectors (PAV)-PAVs contain adenovirus inverted terminal repeats and the minimal adenovirus 5' sequences required for helper virus dependent replication and packaging of the vector. These vectors contain no potentially harmful viral genes, have a theoretical capacity for foreign material of nearly 36 kb, may be produced in reasonably high titers and maintain the tropism of the parent virus for dividing and non-dividing human target cell types.

The PAV vector can be maintained as either a plasmid-borne construct or as an infectious viral particle. As a plasmid construct, PAV is composed of the minimal sequences from wild type adenovirus type 2 necessary for efficient replication and packaging of these sequences and any desired additional exogenous genetic material, by either a wild-type or defective helper virus.

Specifically, PAV contains adenovirus 2 (Ad2) sequences as shown in Figure 17, from nucleotide (nt) 0-356 forming the 5' end of the vector and the last 109 nt of Ad2 forming the 3' end of the construct. The sequences includes the Ad2 flanking inverted terminal repeats (5'ITR) and the 5' ITR adjoining sequences containing the known packaging signal and Ela enhancer. Various convenient restriction sites have been incorporated into the

fragments, allowing the insertion of promoter/gene cassettes which can be packaged in the PAV virion and used for gene transfer (e.g. for gene therapy). The construction and propagation of PAV is described in detail in the following Example 11. By not containing most native adenoviral DNA, the PAVs described herein are less likely to produce a patient immune response or to replicate in a host.

In addition, the PAV vectors can accommodate foreign DNA up to a maximum length of nearly 36 kb. The PAV vectors therefore, are especially useful for cloning larger genes (e.g., CFTR (7.5 kb)); Factor VIII (8 kb); Factor IX (9 kb)), which, traditional vectors have difficulty accommodating. In addition, PAV vectors can be used to transfer more than one gene, or more than one copy of a particular gene. For example, for gene therapy of cystic fibrosis, PAVs can be used to deliver CFTR in conjunction with other genes such as anti proteases (e.g., antiprotease alpha-1-antitrypsin) tissue inhibitor of metalloproteinase, antioxidants (e.g., superoxide dismutase), enhancers of local host defense (e.g., interferons), mucolytics (e.g., DNase); and proteins which block inflammatory cytokines.

#### Ad2-E4/ORF6 Adenovirus Vectors

An adenoviral construct expressing only the open reading frame 6 (ORF6) of adenoviral early region 4 (E4) from the E4 promoter and which is deleted for all other known E4 open reading frames was constructed as described in detail in Example 12. Expression of E4 open reading frame 3 is also sufficient to provide E4 functions required for DNA replication and late protein synthesis. However, it provides these functions with reduced efficiency compared to expression of ORF6, which will likely result in lower levels of virus production. Therefore expressing ORF6, rather than ORF3, appears to be a better choice for producing recombinant adenovirus vectors.

The E4 region of adenovirus is suspected to have a role in viral DNA replication, late mRNA synthesis and host protein synthesis shut off, as well as in viral assembly (Falgout, B. and G. Ketner (1987) *J. Virol.* 61:3759-3768). Adenovirus early region 4 is required for efficient virus particle assembly. Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shutoff. Halbert, D.N. et al. (1985) *J. Virol.* 56:250-257.

The deletion of non-essential open reading frames of E4 increases the cloning capacity of recombinant adenovirus vectors by approximately 2 kb of insert DNA without significantly reducing the viability of the virus in cell culture. When placed in combination with deletions in the E1 and/or E3 regions of adenovirus vectors, the theoretical insert capacity of the resultant vectors is increased to 8-9 kb. An example of where this increased cloning capacity may prove useful is in the development of a gene therapy vector encoding CFTR. As described above, the first generation adenoviral vector approaches the maximum packaging capacity for viral DNA encapsidation. As a result, this virus grows poorly and may occasionally give rise to defective progeny. Including an E4 deletion in the adenovirus

vector should alleviate these problems. In addition, it allows flexibility in the choice of promoters to drive CFTR expression from the virus. For example, strong promoters such as the adenovirus major late promoter, the cytomegalovirus immediate early promoter or a cellular promoter such as the CFTR promoter, which may be too large for first-generation  
5 adenovirus can be used to drive expression.

In addition, by expressing only ORF6 of E4, these second generation adenoviral vectors may be safer for use in gene therapy. Although ORF6 expression is sufficient for viral DNA replication and late protein synthesis in immortalized cells, it has been suggested that ORF6/7 of E4 may also be required in non-dividing primary cells (Hemstrom, C. et al.  
10 (1991) *J. Virol.* 65:1440-1449). The 19 kD protein produced from open reading frame 6 and 7 (ORF6/7) complexes with and activates cellular transcription factor E2F, which is required for maximal activation of early region 2. Early region 2 encodes proteins required for viral DNA replication. Activated transcription factor E2F is present in proliferating cells and is involved in the expression of genes required for cell proliferation (e.g., DHFR, c-myc),  
15 whereas activated E2F is present in lower levels in non-proliferating cells. Therefore, the expression of only ORF6 of E4 should allow the virus to replicate normally in tissue culture cells (e.g., 293 cells), but the absence of ORF6/7 would prevent the potential activation of transcription factor E2F in non-dividing primary cells and thereby reduce the potential for viral DNA replication.

20

#### Target Tissue

Because 95% of CF patients die of lung disease, the lung is a preferred target for gene therapy. The hallmark abnormality of the disease is defective electrolyte transport by the epithelial cells that line the airways. Numerous investigators (reviewed in Quinton, F. (1990)  
25 *FASEB J.* 4:2709) have observed: a) a complete loss of cAMP-mediated transepithelial chloride secretion, and b) a two to three fold increase in the rate of Na<sup>+</sup> absorption. cAMP-stimulated chloride secretion requires a chloride channel in the apical membrane (Welsh, M.J. (1987) *Physiol Rev.* 67:1143-1184). The discovery that CFTR is a phosphorylation-regulated chloride channel and that the properties of the CFTR chloride channel are the same as those  
30 of the chloride channels in the apical membrane, indicate that CFTR itself mediates transepithelial chloride secretion. This conclusion was supported by studies localizing CFTR in lung tissue: CFTR is located in the apical membrane of airway epithelial cells (Denning, G.M. et al. (1992) *J. Cell Biol.* 118:551) and has been reported to be present in the submucosal glands (Taussig et al., (1973) *J. Clin. Invest.* 89:339). As a consequence of loss  
35 of CFTR function, there is a loss of cAMP-regulated transepithelial chloride secretion. At this time it is uncertain how dysfunction of CFTR produces an increase in the rate of Na<sup>+</sup> absorption. However, it is thought that the defective chloride secretion and increased Na<sup>+</sup> absorption lead to an alteration of the respiratory tract fluid and hence, to defective mucociliary clearance, a normal pulmonary defense mechanism. As a result, clearance of

inhaled material from the lung is impaired and repeated infections ensue. Although the presumed abnormalities in respiratory tract fluid and mucociliary clearance provide a plausible explanation for the disease, a precise understanding of the pathogenesis is still lacking.

5           Correction of the genetic defect in the airway epithelial cells is likely to reverse the CF pulmonary phenotype. The identity of the specific cells in the airway epithelium that express CFTR cannot be accurately determined by immunocytochemical means, because of the low abundance of protein. However, functional studies suggest that the ciliated epithelial cells and perhaps nonciliated cells of the surface epithelium are among the main cell types  
10       involved in electrolyte transport. Thus, in practical terms, the present preferred target cell for gene therapy would appear to be the mature cells that line the pulmonary airways. These are not rapidly dividing cells; rather, most of them are nonproliferating and many may be terminally differentiated. The identification of the progenitor cells in the airway is uncertain. Although CFTR may also be present in submucosal glands (Trezise, A.E. and Buchwald, M.  
15       (1991) *Nature* 353:434; Englehardt, J.F. et al. (1992) *J. Clin. Invest.* 90:2598-2607), there is no data as to its function at that site; furthermore, such glands appear to be relatively inaccessible.

          The airway epithelium provides two main advantages for gene therapy. First, access to the airway epithelium can be relatively noninvasive. This is a significant advantage in the  
20       development of delivery strategies and it will allow investigators to monitor the therapeutic response. Second, the epithelium forms a barrier between the airway lumen and the interstitium. Thus, application of the vector to the lumen will allow access to the target cell yet, at least to some extent, limit movement through the epithelial barrier to the interstitium and from there to the rest of the body.

25

#### Efficiency of Gene Delivery Required to Correct The Genetic Defect

          It is unlikely that any gene therapy protocol will correct 100% of the cells that normally express CFTR. However, several observations suggest that correction of a small  
30       percent of the involved cells or expression of a fraction of the normal amount of CFTR may be of therapeutic benefit.

a.       CF is an autosomal recessive disease and heterozygotes have no lung disease. Thus, 50% of wild-type CFTR would appear sufficient for normal function.

35       b.       This issue was tested in mixing experiments using CF cells and recombinant CF cells expressing wild-type CFTR (Johnson, L.G. et al. (1992) *Nature Gen.* 2:21). The data obtained showed that when an epithelium is reconstituted with as few as 6-10% of corrected cells, chloride secretion is comparable to that observed with an epithelium containing 100% corrected cells. Although CFTR expression in the recombinant cells is

probably higher than in normal cells, this result suggests that *in vivo* correction of all CF airway cells may not be required.

5           c.       Recent observations show that CFTR containing some CF-associated mutations retains residual chloride channel activity (Sheppard, D.N. et al. (1992) *Pediatr. Pulmon Suppl.* 8:250; Strong, T.V. et al. (1991) *N. Eng. J. Med.* 325:1630). These mutations are associated with mild lung disease. Thus, even a very low level of CFTR activity may at least partly ameliorate the electrolyte transport abnormalities.

10           d.       As indicated in experiments described below in Example 8, complementation of CF epithelia, under conditions that probably would not cause expression of CFTR in every cell, restored cAMP stimulated chloride secretion.

15           e.       Levels of CFTR in normal human airway epithelia are very low and are barely detectable. It has not been detected using routine biochemical techniques such as immunoprecipitation or immunoblotting and has been exceedingly difficult to detect with immunocytochemical techniques (Denning, G.M. et al. (1992) *J. Cell Biol.* 118:551). Although CFTR has been detected in some cases using laser-scanning confocal microscopy, the signal is at the limits of detection and cannot be detected above background in every case.  
20       Despite that minimal levels of CFTR, this small amount is sufficient to generate substantial cAMP-stimulated chloride secretion. The reason that a very small number of CFTR chloride channels can support a large chloride secretory rate is that a large number of ions can pass through a single channel ( $10^6$ - $10^7$  ions/sec) (Hille, B. (1984) Sinauer Assoc. Inc., Sunderland, MA 420-426).

25

          f.       Previous studies using quantitative PCR have reported that the airway epithelial cells contain at most one to two transcripts per cell (Trapnell, B.C. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:6565).

30           Gene therapy for CF would appear to have a wide therapeutic index. Just as partial expression may be of therapeutic value, overexpression of wild-type CFTR appears unlikely to cause significant problems. This conclusion is based on both theoretical considerations and experimental results. Because CFTR is a regulated channel, and because it has a specific function in epithelia, it is unlikely that overexpression of CFTR will lead to uncontrolled  
35       chloride secretion. First, secretion would require activation of CFTR by cAMP-dependent phosphorylation. Activation of this kinase is a highly regulated process. Second, even if CFTR chloride channels open in the apical membrane, secretion will not ensue without regulation of the basolateral membrane transporters that are required for chloride to enter the cell from the interstitial space. At the basolateral membrane, the sodium-potassium-chloride

cotransporter and potassium channels serve as important regulators of transepithelial secretion (Welsh, M.J. (1987) *Physiol. Rev.* 67:1143-1184).

Human CFTR has been expressed in transgenic mice under the control of the surfactant protein C (SPC) gene promoter (Whitesett, J.A. et al. (1992) *Nature Gen.* 2:13) and the casein promoter (Ditullio, P. et al (1992) *Bio/Technology* 10:74 ). In those mice, CFTR was overexpressed in bronchiolar and alveolar epithelial cells and in the mammary glands, respectively. Yet despite the massive overexpression in the transgenic animals, there were no observable morphologic or functional abnormalities. In addition, expression of CFTR in the lungs of cotton rats produced no reported abnormalities (Rosenfeld, M.A. et al. (1992) *Cell* 68:143-155).

The present invention is further illustrated by the following examples which in no way should be construed as being further limiting. The contents of all cited references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

### EXAMPLES

#### Example 1 - Generation of Full Length CFTR cDNAs

Nearly all of the commonly used DNA cloning vectors are based on plasmids containing modified pMB1 replication origins and are present at up to 500 to 700 copies per cell (Sambrook et al. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989). The partial CFTR cDNA clones isolated by Riordan et al. were maintained in such a plasmid. It was postulated that an alternative theory to intrinsic clone instability to explain the apparent inability to recover clones encoding full length CFTR protein using high copy number plasmids, was that it was not possible to clone large segments of the CFTR cDNA at high gene dosage in *E. coli*. Expression of the CFTR or portions of the CFTR from regulatory sequences capable of directing transcription and/or translation in the bacterial host cell might result in inviability of the host cell due to toxicity of the transcript or of the full length CFTR protein or fragments thereof. This inadvertent gene expression could occur from either plasmid regulatory sequences or cryptic regulatory sequences within the recombinant CFTR plasmid which are capable of functioning in *E. coli*. Toxic expression of the CFTR coding sequences would be greatly compounded if a large number of copies of the CFTR cDNA were present in cells because a high copy number plasmid was used. If the product was indeed toxic as postulated, the growth of cells containing full length and correct sequence would be actively disfavored. Based upon this novel hypothesis, the following procedures were undertaken. With reference to Figure 2, partial CFTR clone T16-4.5 was cleaved with restriction enzymes Sph1 and Pst1 and the resulting 3.9 kb restriction fragment containing exons 11 through most of exon 24 (including



an uncharacterized 119 bp insertion reported by Riordan et al. between nucleotides 1716 and 1717), was isolated by agarose gel purification and ligated between the Sph I and Pst I sites of the pMB1 based vector pkk223-3 (Brosius and Holy, (1984) *Proc. Natl. Acad. Sci.* 81:6929). It was hoped that the pMB1 origin contained within this plasmid would allow it and plasmids constructed from it to replicate at 15-20 copies per host *E. coli* cell (Sambrook et al. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989). The resultant plasmid clone was called pkk-4.5.

Partial CFTR clone T11 was cleaved with Eco RI and Hinc II and the 1.9 kb band encoding the first 1786 nucleotides of the CFTR cDNA plus an additional 100 bp of DNA at the 5' end was isolated by agarose gel purification. This restriction fragment was inserted between the Eco RI site and Sma I restriction site of the plasmid Bluescript Sk- (Stratagene, catalogue number 212206), such that the CFTR sequences were now flanked on the upstream (5') side by a Sal I site from the cloning vector. This clone, designated T11-R, was cleaved with Sal I and Sph I and the resultant 1.8 kb band isolated by agarose gel purification.

Plasmid pkk-4.5 was cleaved with Sal I and Sph I and the large fragment was isolated by agarose gel purification. The purified T11-R fragment and pkk-4.5 fragments were ligated to construct pkk-CFTR1. pkk-CFTR1 contains exons 1 through 24 of the CFTR cDNA. It was discovered that this plasmid is stably maintained in *E. coli* cells and confers no measurably disadvantageous growth characteristics upon host cells.

pkk-CFTR1 contains, between nucleotides 1716 and 1717, the 119 bp insert DNA derived from partial cDNA clone T16-4.5 described above. In addition, subsequent sequence analysis of pkk-CFTR1 revealed unreported differences in the coding sequence between that portion of CFTR1 derived from partial cDNA clone T11 and the published CFTR cDNA sequence. These undesired differences included a 1 base-pair deletion at position 995 and a C to T transition at position 1507.

To complete construction of an intact correct CFTR coding sequence without mutations or insertions and with reference to the construction scheme shown in Figure 3, pkk-CFTR1 was cleaved with Xba I and Hpa I, and dephosphorylated with calf intestinal alkaline phosphatase. In addition, to reduce the likelihood of recovering the original clone, the small unwanted Xba I/Hpa I restriction fragment from pKK-CFTR1 was digested with Sph I. T16-1 was cleaved with Xba I and Acc I and the 1.15 kb fragment isolated by agarose gel purification. T16-4.5 was cleaved with Acc I and Hpa I and the 0.65 kb band was also isolated by agarose gel purification. The two agarose gel purified restriction fragments and the dephosphorylated pKK-CFTR1 were ligated to produce pKK-CFTR2. Alternatively, pKK-CFTR2 could have been constructed using corresponding restriction fragments from the partial CFTR cDNA clone C1-1/5. pKK-CFTR2 contains the uninterrupted CFTR protein coding sequence and conferred slow growth upon *E. coli* host cells in which it was inserted, whereas pKK-CFTR1 did not. The origin of replication of pKK-CFTR2 is derived from pMB1 and confers a plasmid copy number of 15-20 copies per host cell.

#### Example 2 - Improving Host Cell Viability

An additional enhancement of host cell viability was accomplished by a further reduction in the copy number of CFTR cDNA per host cell. This was achieved by transferring the CFTR cDNA into the plasmid vector, pSC-3Z. pSC-3Z was constructed using the pSC101 replication origin of the low copy number plasmid pLG338 (Stoker *et al.*, Gene 18, 335 (1982)) and the ampicillin resistance gene and polylinker of pGEM-3Z (available from Promega). pLG338 was cleaved with Sph I and Pvu II and the 2.8 kb fragment containing the replication origin isolated by agarose gel purification. pGEM-3Z was cleaved with Alw NI, the resultant restriction fragment ends treated with T4 DNA polymerase and deoxynucleotide triphosphates, cleaved with Sph I and the 1.9 kb band containing the ampicillin resistance gene and the polylinker was isolated by agarose gel purification. The pLG338 and pGEM-3Z fragments were ligated together to produce the low copy number cloning vector pSC-3Z. pSC-3Z and other plasmids containing pSC101 origins of replication are maintained at approximately five copies per cell (Sambrook *et al.*, *supra*).

With additional reference to Figure 4, pKK-CFTR2 was cleaved with Eco RV, Pst I and Sal I and then passed over a Sephacryl S400 spun column (available from Pharmacia) according to the manufacturer's procedure in order to remove the Sal I to Eco RV restriction fragment which was retained within the column. pSC-3Z was digested with Sma I and Pst I and also passed over a Sephacryl S400 spun column to remove the small Sma I/Pst I restriction fragment which was retained within the column. The column eluted fractions from the pKK-CFTR2 digest and the pSC-3Z digest were mixed and ligated to produce pSC-CFTR2. A map of this plasmid is presented in Figure 5. Host cells containing CFTR cDNAs at this and similar gene dosages grow well and have stably maintained the recombinant plasmid with the full length CFTR coding sequence. In addition, this plasmid contains a bacteriophage T7 RNA polymerase promoter adjacent to the CFTR coding sequence and is therefore convenient for *in vitro* transcription/translation of the CFTR protein. The nucleotide sequence of CFTR coding region from pSC-CFTR2 plasmid is presented in Sequence Listing 1 as SEQ ID NO:1. Significantly, this sequence differs from the previously published (Riordan, J.R. et al. (1989) *Science* 245:1066-1073) CFTR sequence at position 1990, where there is C in place of the reported A. See Gregory, R.J. et al. (1990) *Nature* 347:382-386. *E. coli* host cells containing pSC-CFTR2, internally identified with the number pSC-CFTR2/AG1, have been deposited at the American Type Culture Collection and given the accession number: ATCC 68244.

#### Example 3 - Alternate Method for Improving Host Cell Viability

A second method for enhancing host cell viability comprises disruption of the CFTR protein coding sequence. For this purpose, a synthetic intron was designed for insertion between nucleotides 1716 and 1717 of the CFTR cDNA. This intron is especially

advantageous because of its easily manageable size. Furthermore, it is designed to be efficiently spliced from CFTR primary RNA transcripts when expressed in eukaryotic cells. Four synthetic oligonucleotides were synthesized (1195RG, 1196RG, 1197RG and 1198RG) collectively extending from the Sph I cleavage site at position 1700 to the Hinc II cleavage site at position 1785 and including the additional 83 nucleotides between 1716 and 1717 (see Figure 6). These oligonucleotides were phosphorylated with T4 polynucleotide kinase as described by Sambrook et al., mixed together, heated to 95°C for 5 minutes in the same buffer used during phosphorylation, and allowed to cool to room temperature over several hours to allow annealing of the single stranded oligonucleotides. To insert the synthetic intron into the CFTR coding sequence and with reference to Figures 7A and 7B, a subclone of plasmid T11 was made by cleaving the Sal I site in the polylinker, repairing the recessed ends of the cleaved DNA with deoxynucleotide triphosphates and the large fragment of DNA Polymerase I and religating the DNA. This plasmid was then digested with Eco RV and Nru I and religated. The resulting plasmid T16-Δ5' extended from the Nru I site at position 490 of the CFTR cDNA to the 3' end of clone T16 and contained single sites for Sph I and Hinc II at positions corresponding to nucleotides 1700 and 1785 of the CFTR cDNA. T16-Δ5' plasmid was cleaved with Sph I and Hinc II and the large fragment was isolated by agarose gel purification. The annealed synthetic oligonucleotides were ligated into this vector fragment to generate T16-intron.

T16-intron was then digested with Eco RI and Sma I and the large fragment was isolated by agarose gel purification. T16-4.5 was digested with Eco RI and Sca I and the 790 bp fragment was also isolated by agarose gel purification. The purified T16-intron and T16-4.5 fragments were ligated to produce T16-intron-2. T16-intron-2 contains CFTR cDNA sequences extending from the Nru I site at position 490 to the Sca I site at position 2818, and includes the unique Hpa I site at position 2463 which is not present in T16-1 or T16-intron-1.

T-16-intron-2 was then cleaved with Xba I and Hpa I and the 1800 bp fragment was isolated by agarose gel purification. pKK-CFTR1 was digested with Xba I and Hpa I and the large fragment was also isolated by agarose gel purification and ligated with the fragment derived from T16-intron-2 to yield pKK-CFTR3, shown in Figure 8. The CFTR cDNA within pKK-CFTR3 is identical to that within pSC-CFTR2 and pKK-CFTR2 except for the insertion of the 83 bp intron between nucleotides 1716 and 1717. The insertion of this intron resulted in improved growth characteristics for cells harboring pKK-CFTR3 relative to cells containing the unmodified CFTR cDNA in pKK-CFTR2.

#### Example 4 - In vitro Transcription/Translation

In addition to sequence analysis, the integrity of the CFTR cDNA open reading frame was verified by *in vitro* transcription/translation. This method also provided the initial CFTR protein for identification purposes. 5 micrograms of pSC-CFTR2 plasmid DNA were linearized with Sal I and used to direct the synthesis of CFTR RNA transcripts with T7 RNA

polymerase as described by the supplier (Stratagene). This transcript was extracted with phenol and chloroform and precipitated with ethanol. The transcript was resuspended in 25 microliters of water and varying amounts were added to a reticulocyte lysate *in vitro* translation system (Promega). The reactions were performed as described by the supplier in the presence of canine pancreatic microsomal membranes (Promega), using <sup>35</sup>S-methionine to label newly synthesized proteins. *In vitro* translation products were analysed by discontinuous polyacrylamide gel electrophoresis in the presence of 0.1% SDS with 8% separating gels (Laemmli, U.K. (1970) *Nature* 227:680-685). Before electrophoresis, the *in vitro* translation reactions were denatured with 3% SDS, 8 M urea and 5% 2-mercaptoethanol in 0.65 M Tris-HCl, pH 6.8. Following electrophoresis, the gels were fixed in methanol:acetic acid:water (30:10:60), rinsed with water and impregnated with 1 M sodium salicylate. <sup>35</sup>S labelled proteins were detected by fluorography. A band of approximately 180 kD was detected, consistent with translation of the full length CFTR insert.

15

#### Example 5 - Elimination of Cryptic Regulatory Signals

Analysis of the DNA sequence of the CFTR has revealed the presence of a potential *E. coli* RNA polymerase promoter between nucleotides 748 and 778 which conforms well to the derived consensus sequence for *E. coli* promoters (Reznikoff and McClure, Maximizing Gene Expression, 1, Butterworth Publishers, Stoneham, MA). If this sequence functions as a promoter functions in *E. coli*, it could direct synthesis of potentially toxic partial CFTR polypeptides. Thus, an additional advantageous procedure for maintaining plasmids containing CFTR cDNAs in *E. coli* would be to alter the sequence of this potential promoter such that it will not function in *E. coli*. This may be accomplished without altering the amino acid sequence encoded by the CFTR cDNA. Specifically, plasmids containing complete or partial CFTR cDNA's would be altered by site-directed mutagenesis using synthetic oligonucleotides (Zoller and Smith, (1983) *Methods Enzymol.* 100:468 ). More specifically, altering the nucleotide sequence at position 908 from a T to C and at position 774 from an A to a G effectively eliminates the activity of this promoter sequence without altering the amino acid coding potential of the CFTR open reading frame. Other potential regulatory signals within the CFTR cDNA for transcription and translation could also be advantageously altered and/or deleted by the same method.

Further analysis has identified a sequence extending from nucleotide 908 to 936 which functions efficiently as a transcriptional promoter element in *E. coli* (Gregory, R.J. et al. (1990) *Nature* 347:382-386). Mutation at position 936 is capable of inactivating this promoter and allowing the CFTR cDNA to be stably maintained as a plasmid in *E. coli* (Cheng, S.H. et al. (1990) *Cell* 63:827-834). Specifically position 936 has been altered from a C to a T residue without the amino acid sequence encoded by the cDNA being altered. Other mutations within this regulatory element described in Gregory, R.J. et al. (1990)

*Nature* 347:382-386 could also be used to inactivate the transcriptional promoter activity. Specifically, the sequence from 908 to 913 (TTGTGA) and from 931 to 936 (GAAAAT) could be altered by site directed mutagenesis without altering the amino acid sequence encoded by the cDNA.

5

#### Example 6 - Cloning of CFTR in Alternate Host Systems

Although the CFTR cDNA displays apparent toxicity in *E. coli* cells, other types of host cells may not be affected in this way. Alternative host systems in which the entire CFTR cDNA protein encoding region may be maintained and/or expressed include other  
10 bacterial species and yeast. It is not possible *a priori* to predict which cells might be resistant and which might not. Screening a number of different host/vector combinations is necessary to find a suitable host tolerant of expression of the full length protein or potentially toxic fragments thereof.

15

#### Example 7 - Generation of Adenovirus Vector Encoding CFTR (Ad2/CFTR)

1. DNA preparation - Construction of the recombinant Ad2/CFTR-1 virus (the sequence of which is shown in Table II and as SEQ ID NO:3) was accomplished as follows: The CFTR cDNA was excised from the plasmid pCMV-CFTR-936C using restriction enzymes  
20 SpeI and EclI361. pCMV-CFTR-936C consists of a minimal CFTR cDNA encompassing nucleotides 123-4622 of the published CFTR sequence cloned into the multiple cloning site of pRC/CMV (Invitrogen Corp.) using synthetic linkers. The CFTR cDNA within this plasmid has been completely sequenced. The SpeI/EclI361 restriction fragment contains 47 bp of 5' sequence derived from synthetic linkers and the multiple cloning site of the vector.

25 The CFTR cDNA (the sequence of which is shown as SEQ ID NO:1 and the amino acid sequence encoded by the CFTR cDNA is shown as SEQ ID NO:2) was inserted between the NheI and SnaBI restriction sites of the adenovirus gene transfer vector pBR-Ad2-7. pBR-Ad2-7 is a pBR322 based plasmid containing an approximately 7 kb insert derived from the 5' 10680 bp of Ad2 inserted between the Clal and BamHI sites of pBR322. From this Ad2  
30 fragment, the sequences corresponding to Ad2 nucleotides 546-3497 were deleted and replaced with a 12 bp multiple cloning site containing an NheI site, an MluI site, and a SnaBI site. The construct also contains the 5' inverted terminal repeat and viral packaging signals, the Ela enhancer and promoter, the Elb 3' intron and the 3' untranslated region and polyadenylation sites. The resulting plasmid was called pBR-Ad2-7/CFTR. Its use to  
35 assemble virus is described below.

2. Virus Preparation from DNA - To generate the recombinant Ad2/CFTR-1 adenovirus, the vector pBR-Ad2-7/CFTR was cleaved with BstBI at the site corresponding to the unique BstBI site at 10670 in Ad2. The cleaved plasmid DNA was ligated to BstBI restricted Ad2

DNA. Following ligation, the reaction was used to transfect 293 cells by the calcium phosphate procedure. Approximately 7-8 days following transfection, a single plaque appeared and was used to reinfect a dish of 293 cells. Following development of cytopathic effect (CPE), the medium was removed and saved. Total DNA was prepared from the infected cells and analyzed by restriction analysis with multiple enzymes to verify the integrity of the construct. Viral supernatant was then used to infect 293 cells and upon development of CPE, expression of CFTR was assayed by the protein kinase A (PKA) immunoprecipitation assay (Gregory, R.J. et al. (1990) *Nature* 347:382 ). Following these verification procedures, the virus was further purified by two rounds of plaque purification.

Plaque purified virus was grown into a small seed stock by inoculation at low multiplicities of infection onto 293 cells grown in monolayers in 925 medium supplemented with 10% bovine calf serum. Material at this stage was designated a Research Viral Seed Stock (RVSS) and was used in all preliminary experiments.

3. Virus Host Cell - Ad2/CFTR-1 is propagated in human 293 cells (ATCC CRL 1573). These cells are a human embryonal kidney cell line which were immortalized with sheared fragments of human Ad5 DNA. The 293 cell line expresses adenovirus early region 1 gene products and in consequence, will support the growth of E1 deficient adenoviruses. By analogy with retroviruses, 293 cells could be considered a packaging cell line, but they differ from usual retrovirus lines in that they do not provide missing viral structural proteins, rather, they provide only some missing viral early functions.

Production lots of virus are propagated in 293 cells derived from the Working Cell Bank (WCB). The WCB is in turn derived from the Master Cell Bank (MCB) which was grown up from a fresh vial of cells obtained from ATCC. Because 293 cells are of human origin, they are being tested extensively for the presence of biological agents. The MCB and WCB are being characterized for identity and the absence of adventitious agents by Microbiological Associates, Rockville, MD.

#### 4. Growth of Production Lots of Virus

Production lots of Ad2/CFTR-1 are produced by inoculation of approximately  $5-10 \times 10^7$  pfu of MVSS onto approximately  $1-2 \times 10^7$  Wcb 293 cells grown in a T175 flask containing 25 mls of 925 medium. Inoculation is achieved by direct addition of the virus (approximately 2-5 mls) to each flask. Batches of 50-60 flasks constitute a lot.

Following 40-48 hours incubation at 37°C, the cells are shaken loose from the flask and transferred with medium to a 250 ml centrifuge bottle and spun at 1000 xg. The cell pellet is resuspended in 4 ml phosphate buffered saline containing 0.1 g/l  $\text{CaCl}_2$  and 0.1g/l  $\text{MgCl}_2$  and the cells subjected to cycles of freeze-thaw to release virus. Cellular debris is removed by centrifugation at 1000 xg for 15 min. The supernatant from this centrifugation is layered on top of the CsCl step gradient: 2 ml 1.4g/ml CsCl and 3 ml 1.25g/ml CsCl in 10

mM Tris, 1 mM EDTA (TE) and spun for 1 hour at 35,000 rpm in a Beckman SW41 rotor. Virus is then removed from the interface between the two CsCl layers, mixed with 1.35 g/ml CsCl in TE and then subjected to a 2.5 hour equilibrium centrifugation at 75,000 rpm in a TLN-100 rotor. Virus is removed by puncturing the side of the tube with a hypodermic  
5 needle and gently removing the banded virus. To reduce the CsCl concentration, the sample is dialyzed against 2 changes of 2 liters of phosphate buffered saline with 10% sucrose.

Following this procedure, dialyzed virus is stable at 4°C for several weeks or can be stored for longer periods at -80°C. Aliquots of material for human use will be tested and while awaiting the results of these tests, the remainder will be stored frozen. The tests to be  
10 performed are described below:

#### 5. Structure and Purity of Virus

SDS polyacrylamide gel electrophoresis of purified virions reveals a number of polypeptides, many of which have been characterized. When preparations of virus were  
15 subjected to one or two additional rounds of CsCl centrifugation, the protein profile obtained was indistinguishable. This indicates that additional equilibrium centrifugation does not purify the virus further, and may suggest that even the less intense bands detected in the virus preparations represent minor virion components rather than contaminating proteins. The identity of the protein bands is presently being established by N-terminal sequence analysis.

20

6. Contaminating Materials - The material to be administered to patients will be  $2 \times 10^6$  pfu,  $2 \times 10^7$  pfu and  $5 \times 10^7$  pfu of purified Ad2/CFTR-1. Assuming a minimum particle to pfu ratio of 500, this corresponds to  $1 \times 10^9$ ,  $1 \times 10^{10}$  and  $2.5 \times 10^{10}$  viral particles, these correspond to a dose by mass of 0.25 µg, 2.5µg and 6.25 µg assuming a molecular mass for  
25 adenovirus of  $150 \times 10^6$ .

The origin of the materials from which a production lot of the purified Ad2/CFTR-1 is derived was described in detail above and is illustrated as a flow diagram in Figure 6. All the starting materials from which the purified virus is made (i.e., MCB, and WCB, and the MVSS) will be extensively tested. Further, the growth medium used will be tested and the  
30 serum will be from only approved suppliers who will provide test certificates. In this way, all the components used to generate a production lot will have been characterized. Following growth, the production lot virus will be purified by two rounds of CsCl centrifugation, dialyzed, and tested. A production lot should constitute  $1-5 \times 10^{10}$  pfu Ad2/CFTR-1.

As described above, to detect any contaminating material aliquots of the production  
35 lot will be analyzed by SDS gel electrophoresis and restriction enzyme mapping. However, these tests have limited sensitivity. Indeed, unlike the situation for purified single chain recombinant proteins, it is very difficult to quantitate the purity of the AD2/CFTR-1 using SDS polyacrylamide gel electrophoresis (or similar methods). An alternative is the immunological detection of contaminating proteins (IDCP). Such an assay utilizes antibodies

raised against the proteins purified in a mock purification run. Development of such an assay has not yet been attempted for the CsCl purification scheme for Ad2/CFTR-1. However, initially an IDCP assay developed for the detection of contaminants in recombinant proteins produced in Chinese hamster ovary (CHO) cells will be used. In addition, to hamster

5 proteins, these assays detect bovine serum albumin (BSA), transferrin and IgG heavy and light chain derived from the serum added to the growth medium. Tests using such reagents to examine research batches of Ad2/CFTR-1 by both ELISA and Western blots are in progress.

Other proteins contaminating the virus preparation are likely to be from the 293 cells - that is, of human origin. Human proteins contaminating therapeutic agents derived from

10 human sources are usually not problematic. In this case, however, we plan to test the production lot for transforming factors. Such factors could be activities of contaminating human proteins or of the Ad2/CFTR-1 vector or other contaminating agents. For the test, it is proposed that 10 dishes of Rat 1 cells containing  $2 \times 10^6$  cells (the number of target cells in the patient) with 4 times the highest human dose of Ad2/CFTR-1 ( $2 \times 10^8$  pfu) will be

15 infected. Following infection, the cells will be plated out in agar and examined for the appearance of transformed foci for 2 weeks. Wild type adenovirus will be used as a control.

Nucleic acids and proteins would be expected to be separated from purified virus preparations upon equilibrium density centrifugation. Furthermore, the 293 cells are not expected to contain VL30 sequences. Biologically active nucleic cells should be detected.

20

Example 8 - Preliminary Experiments Testing the Ability of Ad2/ $\beta$ Gal or Ad2/CFTR Virus to Enter Airway Epithelial Cells

a. Hamster Studies

25 Initial studies involving the intratracheal instillation of the Ad- $\beta$ Gal viral vector into Syrian hamsters, which are reported to be permissive for human adenovirus are being performed. The first study, a time course assessment of the pulmonary and systemic acute inflammatory response to a single intratracheal administration of Ad- $\beta$ Gal viral vector, has been completed. In this study, a total of 24 animals distributed among three treatment

30 groups, specifically, 8 vehicle control, 8 low dose virus ( $1 \times 10^{11}$  particles;  $3 \times 10^8$  pfu), and 8 high dose virus ( $1.7 \times 10^{12}$  particles;  $5 \times 10^9$  pfu), were used. Within each treatment group, 2 animals were analyzed at each of four time points after viral vector instillation: 6 hrs, 24 hrs, 48 hrs, and 7 days. At the time of sacrifice of each animal, lung lavage and blood samples were taken for analysis. The lungs were fixed and processed for normal light-level

35 histology. Blood and lavage fluid were evaluated for total leukocyte count and leukocyte differential. As an additional measure of the inflammatory process, lavage fluid was also evaluated for total protein. Following embeddings, sectioning and hematoxylin/eosin staining, lung sections were evaluated for signs of inflammation and airway epithelial damage.



With the small sample size, the data from this preliminary study were not amenable to statistical analyses, however, some general trends could be ascertained. In the peripheral blood samples, total leukocyte counts showed no apparent dose- or time- dependent changes. In the blood leukocyte differential counts, there may have been a minor dose-related  
5 elevation in percent neutrophil at 6 hours; however, data from all other time points showed no elevation in neutrophil percentages. Taken together, these data suggest little or nor systemic inflammatory response to the viral administration.

From the lung lavage, some elevation in total neutrophil counts were observed at the first three time points (6 hr, 24 hr, 48 hr). By seven days, both total and percent neutrophil  
10 values had returned to normal range. The trends in lung lavage protein levels were more difficult to assess due to inter-animal variability; however, no obvious dose- or time- dependent effects were apparent. First, no damage to airway epithelium was observed at any time point or virus dose level. Second, a time- and dose- dependent mild inflammatory response was observed, being maximal at 48 hr in the high virus dose animals. By seven  
15 days, the inflammatory response had completely resolved, such that the lungs from animals in all treatment groups were indistinguishable.

In summary, a mild, transient, pulmonary inflammatory response appears to be associated with the intratracheal administration of the described doses of adenoviral vector in the Syrian Hamster.

20 A second, single intratracheal dose, hamster study has been initiated. This study is designed to assess the possibility of the spread of ineffective viral vectors to organs outside of the lung and the antibody response of the animals to the adenoviral vector. In this study, the three treatment groups (vehicle control, low dose virus, high dose virus) each contained 12 animals. Animals will be evaluated at three time points: 1 day, 7 days, and 1 month. In this  
25 study, viral vector persistence and possible spread will be evaluated by the assessment of the presence of infective virions in numerous organs including lung, gut, heart, liver, spleen, kidney, brain and gonads. Changes in adenoviral antibody titer will be measured in peripheral blood and lung lavage. Additionally, lung lavage, peripheral blood and lung histology will be evaluated as in the previous study.

30

b. Primate studies.

Studies of recombinant adenovirus are also underway in primates. The goal of these studies is to assess the ability of recombinant adenoviral vectors to deliver genes to the respiratory epithelium *in vivo* and to assess the safety of the construct in primates. Initial  
35 studies in primates targeted nasal epithelia as the site of infection because of its similarity to lower airway epithelia, because of its accessibility, and because nasal epithelia was used for the first human studies. The Rhesus monkey (*Macaca mulatta*) has been chosen for studies, because it has a nasal epithelium similar to that of humans.

How expression of CFTR affects the electrolyte transport properties of the nasal epithelium can be studied in patients with cystic fibrosis. But because the primates have normal CFTR function, instead the ability to transfer a reporter gene was assessed. Therefore the Ad- $\beta$ Gal virus was used. The epithelial cell density in the nasal cavity of the Rhesus monkey is estimated to be  $2 \times 10^6$  cells/cm (based on an average nasal epithelial cell diameter of 7  $\mu$ m) and the surface near 25-50 cm<sup>2</sup>. Thus, there are about  $5 \times 10^7$  cells in the nasal epithelium of Rhesus monkey. To focus especially on safety, the higher viral doses (20-200 MOI) were used *in vivo*. Thus doses in the range of  $10^9$ - $10^{10}$  pfu were used.

In the first pilot study the right nostril of Monkey A was infected with Ad- $\beta$ -Gal (~1 ml). This viral preparation was purified by CsCl gradient centrifugation and then by gel filtration chromatography one week later. Adenoviruses are typically stable in CsCl at 4°C for one to two weeks. However, this viral preparation was found to be defective (i.e., it did not produce detectable  $\beta$ -galactosidase activity in the permissive 293 cells). Thus, it was concluded that there was no live viral activity in the material.  $\beta$ -galactosidase activity in nasal epithelial cells from Monkey A was also not detected. Therefore, in the next study, two different preparations of Ad- $\beta$ -Gal virus: one that was purified on a CsCl gradient and then dialyzed against Tris-buffered saline to remove the CsCl, and a crude unpurified one was used. Titers of Ad- $\beta$ -Gal viruses were  $\sim 2 \times 10^{10}$  pfu/ml and  $> 1 \times 10^{13}$  pfu/ml, respectively, and both preparations produced detectable  $\beta$ -galactosidase activity in 293 cells.

Monkeys were anesthetized by intramuscular injection of ketamine (15 mg/kg). One week before administration of virus, the nasal mucosa of each monkey was brushed to establish baseline cell differentials and levels of  $\beta$ -galactosidase. Blood was drawn for baseline determination of cell differentials, blood chemistries, adenovirus antibody titers, and viral cultures. Each monkey was also examined for weight, temperature, appetite, and general health prior to infection.

The entire epithelium of one nasal cavity was used in each monkey. A foley catheter (size 10) was inserted through each nasal cavity into the pharynx, inflated with 2-3 ml of air, and then pulled anteriorly to obtain tight posterior occlusion at the posterior choana. Both nasal cavities were then irrigated with a solution (~5 ml) of 5 mM dithiothreitol plus 0.2 U/ml neuraminidase in phosphate-buffered saline (PBS) for five minutes. This solution was used to dissolve any residual mucus overlaying the epithelia. (It was subsequently found that such treatment is not required.) The washing procedure also allowed the determination of whether the balloons were effectively isolating the nasal cavity. The virus (Ad- $\beta$ -Gal) was then slowly instilled into the right nostril with the posterior balloon inflated. The viral solution remained in contact with the nasal mucosa for 30 minutes. At the end of 30 minutes, the remaining viral solution was removed by suction. The balloons were deflated, the catheters removed, and the monkey allowed to recover from anesthesia. Monkey A received the CsCl-purified virus (~1.5 ml) and Monkey B received the crude virus (~6 ml). (note that this was the second exposure of Monkey A to the recombinant adenovirus).

Both monkeys were followed daily for appearance of the nasal mucosa, conjunctivitis, appetite, activity, and stool consistency. Each monkey was subsequently anesthetized on days 1, 4, 7, 14, and 21 to obtain nasal, pharyngeal, and tracheal cell samples (either by swabs or brushes) as described below. Phlebotomy was performed over the same time course for hematology, ESR, general screen, antibody serology and viral cultures. Stools were collected every week to assess viral cultures.

To obtain nasal epithelial cells from an anesthetized monkey, the nasal mucosa was first impregnated with 5 drops of Afrin (0.05% oxymetazoline hydrochloride, Schering-Plough) and 1 ml of 2% Lidocaine for 5 min. A cytobrush (the kind typically used for Pap smears) was then used to gently rub the mucosa for about 10 seconds. For tracheal brushings, a flexible fiberoptic bronchoscope; a 3 mm cytology brush (Bard) was advanced through the bronchoscope into the trachea, and a small area was brushed for about 10 seconds. This procedure was repeated twice to obtain a total of  $\sim 10^6$  cells/ml. Cells were then collected on slides (approximately  $2 \times 10^4$  cells/slide using a Cytospin 3 (Shandon, PA)) for subsequent staining (see below).

To determine viral efficacy, nasal, pharyngeal, and tracheal cells were stained for  $\beta$ -galactosidase using X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside). Cleavage of X-gal by  $\beta$ -galactosidase produces a blue color that can be seen with light microscopy. The Ad- $\beta$ -gal vector included a nuclear-localization signal (NLS) (from SV40 large T-antigen) at the amino-terminus of the  $\beta$ -galactosidase sequence to direct expression of this protein to the nucleus. Thus, the number of blue nuclei after staining was determined.

RT-PCR (reverse transcriptase-polymerase chain reaction) was also used to determine viral efficacy. This assay indicates the presence of  $\beta$ -galactosidase mRNA in cells obtained by brushings or swabs. PCR primers were used in both the adenovirus sequence and the LacZ sequence to distinguish virally-produced mRNA from endogenous mRNA. PCR was also used to detect the presence of the recombinant adenovirus DNA. Cytospin preparations were used to assess for the presence of virally produced  $\beta$ -galactosidase mRNA in the respiratory epithelial cells using *in-situ* hybridization. This technique has the advantage of being highly specific and will allow assessment which cells are producing the mRNA.

Whether there was any inflammatory response was assessed by visual inspection of the nasal epithelium and by cytological examination of Wright-stained cells (cytospin). The percentage of neutrophils and lymphocytes were compared to that of the control nostril and to the normal values from four control monkeys. Systemic responses by white blood cell counts, sedimentation rate, and fever were also assessed.

Viral replication at each of the time points was assessed by testing for the presence of live virus in the supernatant of the cell suspension from swabs or brushes. Each supernatant was used to infect (at several dilutions) the virus-sensitive 293 cell line. Cytopathic changes in the 293 cells were monitored for 1 week and then the cells were fixed and stained for  $\beta$ -galactosidase. Cytopathic effects and blue-stained cells indicated the presence of live virus.

Positive supernatants will also be subjected to analysis of nonintegrating DNA to identify (confirm) the contributing virus(es).

Antibody titers to type 2 adenovirus and to the recombinant adenovirus were determined by ELISA. Blood/serum analysis was performed using an automated chemistry analyzer Hitachi 737 and an automated hematology analyzer Technicom H6. The blood buffy coat was cultured in A549 cells for wild type adenovirus and was cultured in the permissive 293 cells.

**Results:** Both monkeys tolerated the procedure well. Daily examination revealed no evidence of coryza, conjunctivitis or diarrhea. For both monkeys, the nasal mucosa was mildly erythematous in both the infection side and the control side; this was interpreted as being due to the instrumentation. Appetites and weights were not affected by virus administered in either monkey. Physical examination on days 1, 4, 7, 14 and 21 revealed no evidence of lymphadenopathy, tachypnea, or tachycardia. On day 21, monkey B had a temperature 39.1°C (normal for Rhesus monkey 38.8°C) but had no other abnormalities on physical exam or in laboratory data. Monkey A had a slight leukocytosis on day 1 post infection which returned to normal by day 4; the WBC was 4,920 on the day of infection, 8,070 on day 1, and 5,200 on day 4. The ESR did not change after the infection. Electrolytes and transaminases were normal throughout.

Wright stains of cells from nasal brushing were performed on days 4, 7, 14, and 21. They revealed less than 5% neutrophils and lymphocytes. There was no difference between the infected and the control side.

X-Gal stains of the pharyngeal swabs revealed blue-stained cells in both monkeys on days 4, 7, and 14; only a few of the cells had clear nuclear localization of the pigment and some pigment was seen in extracellular debris. On day 7 post infection, X-Gal stains from the right nostril of monkey A, revealed a total of 135 ciliated cells with nuclear-localized blue stain. The control side had only 4 blue cells. Monkey B had 2 blue cells from the infected nostril and none from the control side. Blue cells were not seen on day 7, 14, or 21.

RT-PCR on day 3 post infection revealed a band of the correct size that hybridized with a  $\beta$ -Gal probe, consistent with  $\beta$ -Gal mRNA in the samples from Monkey A control nostril and Monkey B infected nostril. On day 7 there was a positive band in the sample from the infected nostril of Monkey A, the same specimen that revealed blue cells.

Fluid from each nostril, the pharynx, and trachea of both monkeys was placed on 293 cells to check for the presence of live virus by cytopathic effect and X-Gal stain. In Monkey A, live virus was detected in both nostrils on day 3 after infection; no live virus was detected at either one or two weeks post-infection. In Monkey B, live virus was detected in both nostrils, pharynx, and trachea on day 3, and only in the infected nostril on day 7 after infection. No live virus was detected 2 weeks after the infection.

c. Human Explant Studies

In a second type of experiment, epithelial cells from a nasal polyp of a CF patient were cultured on permeable filter supports. These cells form an electrically tight epithelial monolayer after several days in culture. Eight days after seeding, the cells were exposed to the Ad2/CFTR virus for 6 hours. Three days later, the short-circuit current (I<sub>sc</sub>) across the monolayer was measured. cAMP agonists did not increase the I<sub>sc</sub>, indicating that there was no change in chloride secretion. However, this defect was corrected after infection with recombinant Ad2/CFTR. Cells infected with Ad2/CFTR (MOI=5; MOI refers to multiplicity of infection; 1 MOI indicates one pfu/cell) express functional CFTR; cAMP agonists stimulated I<sub>sc</sub>, indicating stimulation of Cl<sup>-</sup> secretion. Ad2/CFTR also corrected the CF chloride channel defect in CF tracheal epithelial cells. Additional studies indicated that Ad2/CFTR was able to correct the chloride secretory defect without altering the transepithelial electrical resistance; this result indicates that the integrity of the epithelial cells and the tight junctions was not disrupted by infection with Ad2/CFTR. Application of 1 MOI of Ad2/CFTR was also found to be sufficient to correct the CF chloride secretory defect.

The experiments using primary cultures of human airway epithelial cells indicate that the Ad2/CFTR virus is able to enter CF airway epithelial cells and express sufficient CFTR to correct the defect in chloride transport.

20 Example 9 -In Vivo Delivery to and Expression of CFTR in Cotton Rat and Rhesus Monkey Epithelium

**MATERIALS AND METHODS**

Adenovirus vector

25 Ad2/CFTR-1 was prepared as described in Example 7. The DNA construct comprises a full length copy of the Ad2 genome of approximately 37.5 kb from which the early region 1 genes (nucleotides 546 to 3497) have been replaced by cDNA for CFTR (nucleotides 123 to 4622 of the published CFTR sequence with 53 additional linker nucleotides). The viral Ela promoter was used for CFTR cDNA. Termination/polyadenylation occurs at the site normally used by the Elb and protein IX transcripts. The recombinant virus E3 region was conserved. The size of the Ad2-CFTR-1 vector is approximately 104.5% that of wild-type adenovirus. The recombinant virus was grown in 293 cells that complement the E1 early viral promoters. The cells were frozen and thawed three times to release the virus and the preparation was purified on a CsCl gradient, then dialyzed against Tris-buffered saline (TBS) to remove the CsCl, as described.

### Animals

**Rats.** Twenty two cotton rats (6-8 weeks old, weighing between 80-100 g) were used for this study. Rats were anesthetized by inhaled methoxyflurane (Pitman Moore, Inc., Mundelen, Ill). Virus was applied to the lungs by nasal instillation during inspiration.

5 Two cotton rat studies were performed. In the first study, seven rats were assigned to a one time pulmonary infection with 100  $\mu$ l solution containing  $4.1 \times 10^9$  plaque forming units (pfu) of the Ad2/CFTR-1 virus and 3 rats served as controls. One control rat and either two or three experimental rats were sacrificed with methoxyflurane and studies at each of three time points: 4, 11, or 15 days after infection.

10 The second group of rats was used to test the effect of repeat administration of the recombinant virus. All 12 rats received  $2.1 \times 10^8$  pfu of the Ad2/CFTR-1 virus on day 0 and 9 of the rats received a second dose of  $3.2 \times 10^8$  pfu of Ad2/CFTR-1 14 days later. Groups of one control rat and three experimental rats were sacrificed at 3, 7, or 14 days after the second administration of virus. Before necropsy, the trachea was cannulated and  
15 bronchoalveolar lavage (BAL) was performed with 3 ml aliquots of phosphate-buffered saline. A median sternotomy was performed and the right ventricle cannulated for blood collection. The right lung and trachea were fixed in 4% formaldehyde and the left lung was frozen in liquid nitrogen and kept at  $-70^\circ\text{C}$  for evaluation by immunochemistry, reverse transcriptase polymerase chain reaction (RT-PCR), and viral culture. Other organs were removed and  
20 quickly frozen in liquid nitrogen for evaluation by polymerase chain reaction (PCR).

**Monkeys.** Three female Rhesus monkeys were used for this study; a fourth female monkey was kept in the same room, and was used as control. For application of the virus, the monkeys were anesthetized by intramuscular injection of ketamine (15 mg/kg). The entire epithelium of one nasal cavity in each monkey was used for virus application. A foley  
25 catheter (size 10) was inserted through each nasal cavity into the pharynx, the balloon was inflated with 2-3 ml of air, and then pulled anteriorly to obtain a tight occlusion at the posterior choana. The Ad2/CFTR-1 virus was then instilled slowly in the right nostril with the posterior balloon inflated. The viral solution remained in contact with the nasal mucosa for 30 min. The balloons were deflated, the catheters were removed, and the monkeys were  
30 allowed to recover from anesthesia. A similar procedure was performed on the left nostril, except that TBS solution was instilled as a control. The monkeys received a total of three doses of the virus over a period of 5 months. The total dose given was  $2.5 \times 10^9$  pfu the first time,  $2.3 \times 10^9$  pfu the second time, and  $2.8 \times 10^9$  pfu the third time. It was estimated that the cell density of the nasal epithelia to be  $2 \times 10^6$  cells/cm<sup>2</sup> and a surface area of 25 to 50  
35 cm<sup>2</sup>. This corresponds to a multiplicity of infection (MOI) of approximately 25.

The animals were evaluated 1 week before the first administration of virus, on the day of administration, and on days 1, 3, 6, 13, 21, 27, and 42 days after infection. The second administration of virus occurred on day 55. The monkeys were evaluated on day 55 and then on days 56, 59, 62, 69, 76, 83, 89, 96, 103, and 111. For the third administration, on day 134,

only the left nostril was cannulated and exposed to the virus. The control monkey received instillations of PBS instead of virus. Biopsies of the left medial turbinate were carried out on day 135 in one of the infected monkeys, on day 138 on the second infected monkey, and on day 142 on the third infected monkey and on the control monkey.

- 5 For evaluations, monkeys were anesthetized by intramuscular injection of ketamine (15 mg/kg). To obtain nasal epithelial cells, the nasal mucosa was first impregnated with 5 drops of Afrin (0.05% oxymetazoline hydrochloride, Schering-Plough) and 1 ml of 2% Lidocaine for 5 minutes. A cytobrush was then used to gently rub the mucosa for about 3 sec. To obtain pharyngeal epithelial swabs, a cotton-tipped applicator was rubbed over the  
10 back of the pharynx 2-3 times. The resulting cells were dislodged from brushes or applicators into 2 ml of sterile PBS. Biopsies of the medial turbinate were performed using cupped forceps under direct endoscopic control.

- Animals were evaluated daily for evidence of abnormal behavior of physical signs. A record of food and fluid intake was used to assess appetite and general health. Stool  
15 consistency was also recorded to check for the possibility of diarrhea. At each of the evaluation time points, rectal temperature, respiratory rate, and heart rate were measured. The nasal mucosa, conjunctivas, and pharynx were visually inspected. The monkeys were also examined for lymphadenopathy.

- Venous blood from the monkeys was collected by standard venipuncture technique.  
20 Blood/serum analysis was performed in the clinical laboratory of the University of Iowa Hospitals and Clinics using a Hitachi 737 automated chemistry analyzer and a Technicom H6 automated hematology analyzer.

#### Serology

- 25 Sera were obtained and anti-adenoviral antibody titers were measured by an enzyme-linked immunoadsorbant assay (ELISA). For the ELISA, 50 ng/well of filled adenovirus (Lee Biomolecular Research Laboratories, San Diego, Ca) in 0.1M NaHCO<sub>3</sub> were coated on 96 well plates at 4°C overnight. The test samples at appropriate dilutions were added, starting at a dilution of 1/50. The samples were incubated for 1 hour, the plates washed, and  
30 a goat anti-human IgG HRP conjugate (Jackson ImmunoResearch Laboratories, West Grove, PA) was added and incubated for 1 hour. The plates were washed and O-Phenylenediamine (Sigma Chemical Co., St. Louis, MO) was added for 30 min. at room temperature. The assay was stopped with 4.5 M H<sub>2</sub>SO<sub>4</sub> and read at 490 nm on a Molecular Devices microplate reader. The titer was calculated as the product of the reciprocal of the initial dilution and the  
35 reciprocal of the dilution in the last well with an OD>0.100.

Neutralizing antibodies measure the ability of the monkey serum to prevent infection of 293 cells by adenovirus. Monkey serum (1:25 dilution) [or nasal washings (1:2 dilutions)] was added in two-fold serial dilutions to a 96 well plate. Adenovirus (2.5 x 10<sup>5</sup> pfu) was added and incubated for 1 hour at 37°C. The 293 cells were then added to all wells and the

plates were incubated until the serum-free control wells exhibited >95% cytopathic effect. The titer was calculated as the product of the reciprocal of the initial dilution times the reciprocal of the dilution in the last well showing >95% cytopathic effect.

5 Bronchoalveolar lavage and nasal brushings for cytology

Bronchoalveolar lavage (BAL) was performed by cannulating the trachea with a silastic catheter and injecting 5 ml of PBS. Gentle suction was applied to recover the fluid. The BAL sample was spun at 5000 rpm for 5 min. and cells were resuspended in 293 media at a concentration of  $10^6$  cells/ml. Cells were obtained from the monkey's nasal epithelium by gently rubbing the nasal mucosa for about 3 sec. with a cytobrush. The resulting cells were dislodged from the brushes into 2 ml of PBS. Forty microliters of the cell suspension were cytocentrifuged onto slides and stained with Wright's stain. Samples were examined by light microscopy.

15

Histology of lung sections and nasal biopsies

The right lung of each cotton rat was removed, inflated with 4% formaldehyde, and embedded in paraffin for sectioning. Nasal biopsies from the monkeys were also fixed with 4% formaldehyde. Histologic sections were stained with hematoxylin and eosin (H&E). Sections were reviewed by at least one of the study personnel and by a pathologist who was unaware of the treatment each rat received.

Immunocytochemistry

Pieces of lung and trachea of the cotton rats and nasal biopsies were frozen in liquid nitrogen on O.C.T. compound. Cryosections and paraffin sections of the specimens were used for immunofluorescence microscopy. Cytospin slides of nasal brushings were prepared on gelatin coated slides and fixed with paraformaldehyde. The tissue was permeabilized with Triton X-100, then a pool of monoclonal antibodies to CFTR (M13-1, M1-4) (Denning, G.M. et al. (1992) *J. Clin. Invest.* 89:339-349) was added and incubated for 12 hours. The primary antibody was removed and an anti-mouse biotinylated antibody (Biomed, Foster City, CA) was added. After removal of the secondary antibody, streptavidin FITC (Biomed, Foster City, Ca) was added and the slides were observed under a laser scanning confocal microscope. Both control animal samples and non-immune IgG stained samples were used as controls.

35

PCR

PCR was performed on pieces of small bowel, brain, heart, kidney, liver, ovaries, and spleen from cotton rats. Approximately 1 g of the rat organs was mechanically ground and mixed with 50  $\mu$ l sterile water, boiled for 5 min., and centrifuged. A 5  $\mu$ l aliquot of the



supernatant was removed for further analysis. Monkey nasal brushings suspensions were also used for PCR.

- Nested PCR primer sets were designed to selectively amplify Ad2/CFTR-1 DNA over endogenous CFTR by placing one primer from each set in the adenovirus sequence and the other primer in the CFTR sequence. The first primer set amplifies a 723 bp fragment and is shown below:

Ad2 5' ACT CTT GAG TGC CAG CGA GTA GAG TTT TCT CCT CCG 3' (SEQ ID NO:4)

CFTR 5' GCA AAG GAG CGA TCC ACA CGA AAT GTG CC 3' (SEQ ID NO:5)

- The nested primer set amplifies a 506 bp fragment and is shown below:

Ad2 5' CTC CTC CGA GCC GCT CCG AGC TAG 3' (SEQ ID NO:6)

CFTR 5' CCA AAA ATG GCT GGG TGT AGG AGC AGT GTC C 3' (SEQ ID NO:7)

- A PCR reaction mix containing 10mM Tris-Cl (pH 8.3), 50mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin, 400 µM each dNTP, 0.6 µM each primer (first set), and 2.5 units AmpliTaq (Perkin Elmer) was aliquoted into separate tubes. A 5 µl aliquot of each sample prep was then added and the mixture was overlaid with 50 µl of light mineral oil. The samples were processed on a Barnstead/Thermolyne (Dubuque, IA) thermal cycler programmed for 1 min. at 94°C, 1 min. at 65°C, and 2 min. at 72°C for 40 cycles. Post-run dwell was for 7 min. at 72°C. A 5 µl aliquot was removed and added to a second PCR reaction using the nested set of primers and cycled as above. A 10 µl aliquot of the final amplification reaction was analyzed on a 1% agarose gel and visualized with ethidium bromide.

- To determine the sensitivity of this procedure, a PCR mix containing control rat liver supernatant was aliquoted into several tubes and spiked with dilutions of Ad2/CFTR-1. Following the amplification protocols described above, it was determined that the nested PCR procedure could detect as little as 50 pfu of viral DNA.

#### RT-PCR

- RT-PCR was used to detect vector-generated mRNA in cotton rat lung tissue and samples from nasal brushings from monkeys. A 200 µl aliquot of guanidine isothiocyanate solution (4 M guanidine isothiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, and 0.1 M β-mercaptoethanol) was added to a frozen section of each lung and pellet from nasal brushings and the tissue was mechanically ground. Total RNA was isolated utilizing a single-step method (Chomczynski, P. and Sacchi, N. et al. (1987) *Analytical Biochemistry* 162:156-159; Hanson, C.A. et al. (1990) *Am. J. Pathol.* 137:1-6). The RNA was incubated with 1 unit RQ1 RNase-free DNase (Promega Corp., Madison WI) at 37°C for 20 min., denatured at 99°C for 5 min., precipitated with ammonium acetate and ethanol, and redissolved in 4 µl diethylpyrocarbonate treated water containing 20 units RNase Block 1 (Stratagene, La Jolla CA). A 2 µl aliquot of the purified RNA was reverse transcribed using

the GeneAmp RNA PCR kit (Perkin Elmer Cetus) and the downstream primer from the first primer set described in the previous section. Reverse transcriptase was omitted from the reaction with the remaining 2  $\mu$ l of the purified RNA prep, as a control in which preparations (both +/- RT) were then amplified using nested primer sets and the PCR protocols described above. A 10  $\mu$ l aliquot of the final amplification reaction was analyzed on a 1% agarose gel and visualized with ethidium bromide.

#### Southern analysis.

To verify the identity of the PCR products, Southern analysis was performed. The DNA was transferred to a nylon membrane as described (Sambrook *et al.*, *supra*). A fragment of CFTR cDNA (amino acids #1-525) was labeled with [<sup>32</sup>P]-dCTP (ICN Biomedicals, Inc. Irvine CA) using an oligolabeling kit (Pharmacia, Piscataway, NJ) and purified over a NICK column (Pharmacia Piscataway, NJ) for use as a hybridization probe. The labeled probe was denatured, cooled, and incubated with the prehybridized filter for 15 hours at 42°C. The hybridized filter was then exposed to film (Kodak XAR-5) for 10 min.

#### Culture of Ad2/CFTR-1

Viral cultures were performed on the permissive 293 cell line. For culture of virus from lung tissue, 1 g of lung was frozen/thawed 3-6 times and then mechanically disrupted in 200  $\mu$ l of 293 media. For culture of BAL and monkey nasal brushings, the cell suspension was spun for 5 min and the supernatant was collected. Fifty  $\mu$ l of the supernatant was added in duplicate to 293 cells grown in 96 well plates at 50% confluence. The 293 cells were incubated for 72 hr at 37°C, then fixed with a mixture of equal parts of methanol and acetone for 10 min. and incubated with FITC-labeled anti-adenovirus monoclonal antibodies (Chemicon, Light Diagnostics, Temecuca, CA) for 30 min. Positive nuclear immunofluorescence was interpreted as positive culture. The sensitivity of the assay was evaluated by adding dilutions of Ad2/CFTR-1 to 50  $\mu$ l of the lung homogenate from one of the control rats. Viral replication was detected when as little as 1 pfu was added.

## **RESULTS**

#### Efficacy of Ad2/CFTR-1 in the lungs of cotton rats.

To test the ability of Ad2/CFTR-1 to transfer CFTR cDNA to the intrapulmonary airway epithelium, several studies were performed. 4 x 10<sup>6</sup> pfu - IU of Ad2/CFTR-1 in 100  $\mu$ l was administered to seven cotton rats; three control rats received 100  $\mu$ l of TBS (the vehicle for the virus). The rats were sacrificed 4, 10 or 14 days later. To detect viral transcripts encoding CFTR, reverse transcriptase was used to prepare cDNA from lung homogenates. The cDNA was amplified with PCR using primers that span adenovirus and CFTR-encoded

sequences. Thus, the procedure did not detect endogenous rat CFTR. Figure 16 shows that the lungs of animals which received Ad2/CFTR-1 were positive for virally-encoded CFTR mRNA. The lungs of all control rats were negative.

To detect the protein, lung sections were immunostained with antibodies specific to CFTR. CFTR was detected at the apical membrane of bronchial epithelium from all rats exposed to Ad2/CFTR-1, but not from control rats. The location of recombinant CFTR at the apical membrane is consistent with the location of endogenous CFTR in human airway epithelium. Recombinant CFTR was detected above background levels because endogenous levels of CFTR in airway epithelia are very low and thus, difficult to detect by immunocytochemistry (Trapnell, B. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:6565-6569; Denning, G.M. et al. (1992) *J. Cell Biol.* 118:551-59).

These results show that Ad2/CFTR-1 directs the expression of CFTR mRNA in the lung of the cotton rat and CFTR protein in the intrapulmonary airways.

#### 15 Safety of Ad2/CFTR-1 in cotton rats.

Because the E1 region of Ad2 is deleted in the Ad2/CFTR-1 virus, the vector was expected to be replication-impaired (Berkner, K.L. (1988) *BioTechniques* 6:616-629) and that it would be unable to shut off host cell protein synthesis (Basuss, L.E. et al. (1989) *J. Virol.* 50:202-212). Previous *in vitro* studies have suggested that this is the case in a variety of cells including primary cultures of human airway epithelial cells (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476). However, it is important to confirm this *in vivo* in the cotton rat, which is the most permissive animal model for human adenovirus infection (Ginsberg, H.S. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3823-3827; Prince, G.A. et al. (1993) *J. Virol.* 67:101-111). Although dose of virus of  $4.1 \times 10^{10}$  pfus per kg was used, none of the rats died. More importantly, extracts from lung homogenates from each of the cotton rats were cultured in the permissive 293 cell line. With this assay 1 pfu of recombinant virus was detected in lung homogenate. However, virus was not detected by culture in the lungs of any of the treated animals. Thus, the virus did not appear to replicate *in vivo*.

It is also possible that administration of Ad2/CFTR-1 could cause an inflammatory response, either due to a direct effect of the virus or as a result of administration of viral particles. Several studies were performed to test this possibility. None of the rats had a change in the total or differential white blood cell count, suggesting that there was no major systemic inflammatory response. To assess the pulmonary inflammatory response more directly, bronchoalveolar lavage was performed on each of the rats (Figures 17A and 17B). Figure 17A shows that there was no change in the total number of cells recovered from the lavage or in the differential cell count.

Sections of the lung stained by H&E were also prepared. There was no evidence of viral inclusions or any other changes characteristic of adenoviral infection (Prince, G.A. et al. (1993) *J. Virol.* 67:101-111). When coded lung sections were evaluated by a skilled reader

who was unaware of which sections were treated, she was unable to distinguish between sections from the treated and untreated lungs.

It seemed possible that the recombinant adenovirus could escape from the lung into other tissues. To test for this possibility, other organs from the rats were evaluated using nested PCR to detect viral DNA. All organs tested from infected rats were negative, with the exception of small bowel which was positive in 3 of 7 rats. Figure 18 shows the results of 2 infected rats and one control rat sacrificed on day 4 after infection. The organ homogenates from the infected rats sacrificed were negative for Ad2/CFTR-1 with the exception of the small bowel. Organ homogenates from control rats sacrificed on day 4 after infection were negative for Ad2/CFTR-1. The presence of viral DNA in the small bowel suggests that the rats may have swallowed some of the virus at the time of instillation or, alternatively, the normal airway clearance mechanisms may have resulted in deposition of viral DNA in the gastrointestinal tract. Despite the presence of viral DNA in homogenates of small intestine, none of the rats developed diarrhea. This result suggests that if the virus expressed CFTR in the intestinal epithelium, there was no obvious adverse consequence.

#### Repeat administration of Ad2/CFTR-1 to cotton rats

Because adenovirus DNA integration into chromosomal DNA is not necessary for gene expression and only occurs at very low frequency, expression following any given treatment was anticipated to be finite and that repeated administration of recombinant adenovirus would be required for treatment of CF airway disease. Therefore, the effect of repeated administration of Ad2/CFTR-1 cotton rats was examined. Twelve cotton rats received 50 µl of Ad2/CFTR-1. Two weeks later, 9 of the rats received a second dose of 50 µl of Ad2/CFTR-1 and 3 rats received 50 µl of TBS. Rats were sacrificed on day 3, 7, or 14 after virus administration. At the time of the second vector administration all cotton rats had an increased antibody titer to adenovirus.

After the second intrapulmonary administration of virus, none of the rats died. Moreover, the results of studies assessing safety and efficacy were similar to results obtained in animals receiving adenovirus for the first time. Viral cultures of rat lung homogenates on 293 cells were negative at all time points, suggesting that there was no virus replication. There was no difference between treated and control rats in the total or differential white blood count at any of the time points. The lungs were evaluated by histologic sections stained with H&E; and found no observable differences between the control and treated rats when sections were read by us or by a blinded skilled reader. Examples of some sections are shown in Figure 19. When organs were examined for viral DNA using PCR, viral DNA was found only in the small intestine of 2 rats. Despite seropositivity of the rats at the time of the second administration, expression of CFTR (as assessed by RT-PCR and by immunocytochemistry of sections stained with CFTR antibodies) similar to that seen in animals that received a single administration was observed.

These results suggest that prior administration of Ad2/CFTR-1 and the development of an antibody response did not cause an inflammatory response in the rats nor did it prevent virus-dependent production of CFTR.

5 Evidence that Ad2/CFTR-1 expresses CFTR in primate airway epithelium

The cells lining the respiratory tract and the immune system of primates are similar to those of humans. To test the ability of Ad2/CFTR-1 to transfer CFTR to the respiratory epithelium of primates, Ad2/CFTR was applied on three occasions as described in the methods to the nasal epithelium of three Rhesus monkeys. To obtain cells from the  
10 respiratory epithelium, the epithelium was brushed using a procedure similar to that used to sample the airway epithelium of humans during fiberoptic bronchoscopy.

To assess gene transfer, RT-PCR was used as described above for the cotton rats. RT-PCR was positive on cells brushed from the right nostril of all three monkeys, although it was only detectable for 18 days after virus administration. An example of the results are  
15 shown in Figure 20A. The presence of a positive reaction in cells from the left nostril most likely represents some virus movement to the left side due to drainage, or possibly from the monkey moving the virus from one nostril to the other with its fingers after it recovered from anesthesia.

The specificity of the RT-PCR is shown in Figure 20B. A Southern blot with a probe  
20 to CFTR hybridized with the RT-PCR product from the monkey infected with Ad2/CFTR-1. As a control, one monkey received a different virus (Ad2/ $\beta$ Gal-1) which encodes  $\beta$ -galactosidase. When different primers were used to reverse transcribe the  $\beta$ -galactosidase mRNA and amplify the cDNA, the appropriate PCR product was detected. However, the PCR product did not hybridize to the CFTR probe on Southern blot. This result shows the  
25 specificity of the reaction for amplification of the adenovirus-directed CFTR transcript.

The failure to detect evidence of adenovirus-encoded CFTR mRNA at 18 days or beyond suggests that the sensitivity of the RT-PCR may be low because of limited efficacy of the reverse transcriptase or because RNases may have degraded RNA after cell acquisition. Viral DNA, however, was detected by PCR in brushings from the nasal epithelium for  
30 seventy days after application of the virus. This result indicates that although mRNA was not detected after 2 weeks, viral DNA was present for a prolonged period and may have been transcriptionally active.

To assess the presence of CFTR proteins directly, cells obtained by brushing were plated onto slides by cytospin and stained with antibodies to CFTR. Figure 21 shows an  
35 example of the immunocytochemistry of the brushed cells. A positive reaction is clearly evident in cells exposed to Ad2/CFTR-1. The cells were scored as positive by immunocytochemistry when evaluated by a reader uninformed to the identity of the samples. Immunocytochemistry remained positive for five to six weeks for the three monkeys, even after the second administration of Ad2/CFTR-1. On occasion, a few positive staining cells

were observed from the contralateral nostril of the monkeys. However, this was of short duration, lasting at most one week.

Sections of nasal turbinate biopsies obtained within a week after the third infection were also examined. In sections from the control monkey, little if any immunofluorescence from the surface epithelium was observed, but the submucosal glands showed significant staining of CFTR (Fig. 22). These observations are consistent with results of previous studies (Engelhardt, J.F. and Wilson, J.M. (1992) *Nature Gen.* 2:240-248.) In contrast, sections from monkeys that received Ad2/CFTR-1 revealed increased immunofluorescence at the apical membrane of the surface epithelium. The submucosal glands did not appear to have greater immunostaining than was observed under control conditions. These results indicate that Ad2/CFTR-1 can transfer the CFTR cDNA to the airway epithelium of Rhesus monkeys, even in seropositive animals (see below).

#### Safety of Ad2/CFTR-1 administered to monkeys

Figure 23 shows that all three treated monkeys developed antibodies against adenovirus. Antibody titers measured by ELISA rose within two weeks after the first infection. With subsequent infections the titer rose within days. The sentinel monkey had low antibody titers throughout the experiment. Tests for the presence of neutralizing antibodies were also performed. After the first administration, neutralizing antibodies were not observed, but they were detected after the second administration and during the third viral administration (Fig. 23).

To detect virus, supernatants from nasal brushings and swabs were cultured on 293 cells. All monkeys had positive cultures on day 1 and on day 3 or 4 from the infected nostril. Cultures remained positive in one of the monkeys at seven days after administration, but cultures were never positive beyond 7 days. Live virus was occasionally detected in swabs from the contra lateral nostril during the first 4 days after infection. The rapid loss of detectable virus suggests that there was not viral replication. Stools were routinely cultured, but virus was never detected in stools from any of the monkeys.

None of the monkeys developed any clinical signs of viral infection or inflammation. Visual inspection of the nasal epithelium revealed slight erythema in all three monkeys in both nostrils on the first day after infection; but similar erythema was observed in the control monkey and likely resulted from the instrumentation. There was no visible abnormalities at days 3 or 4, or on weekly inspection thereafter. Physical examination revealed no fever, lymphadenopathy, conjunctivitis, tachypnea, or tachycardia at any of the time points. No abnormalities were found in a complete blood count or sedimentation rate, nor were abnormalities observed in serum electrolytes, transaminases, or blood urea nitrogen and creatinine.

Examination of Wright-stained cells from the nasal brushings showed that neutrophils and lymphocytes accounted for less than 5% of total cells in all three monkeys.

Administration of the Ad2/CFTR-1 caused no change in the distribution or number of inflammatory cells at any of the time points following virus administration. H&E stains of the nasal turbinate biopsies specimens from the control monkey could not be differentiated from that of the experimental monkey when the specimens were reviewed by an independent pathologist. (Fig. 24)

These results demonstrate the ability of a recombinant adenovirus encoding CFTR (Ad2/CFTR-1) to express CFTR cDNA in the airway epithelium of cotton rats and monkeys during repeated administration. They also indicate that application of the virus involves little if any risk. Thus, they suggest that such a vector may be of value in expressing CFTR in the airway epithelium of humans with cystic fibrosis.

Two methods were used to show that Ad2/CFTR-1 expresses CFTR in the airway epithelium of cotton rats and primates: CFTR mRNA was detected using RT-PCR and protein was detected by immunocytochemistry. Duration of expression as assessed immunocytochemically was five to six weeks. Because very little protein is required to generate  $\text{Cl}^-$  secretion (Welsh, M.J. (1987) *Physiol. Rev.* 67:1143-1184; Trapnell, B.C. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:6565-6569; Denning, G.M. et al. (1992) *J. Cell Biol.* 118:551-559), it is likely that functional expression of CFTR persists substantially longer than the period of time during which CFTR was detected by immunocytochemistry. Support for this evidence comes from two considerations: first, it is very difficult to detect CFTR immunocytochemically in the airway epithelium, yet the expression of an apical membrane  $\text{Cl}^-$  permeability due to the presence of CFTR  $\text{Cl}^-$  channels is readily detected. The ability of a minimal amount of CFTR to have important functional effects is likely a result of the fact that a single ion channel conducts a very large number of ions ( $10^6$  -  $10^7$  ions/sec). Thus, ion channels are not usually abundant proteins in epithelia. Second, previous work suggests that the defective electrolyte transport of CF epithelia can be corrected when only 6-10% of cells in a CF airway epithelium overexpress wild-type CFTR (Johnson, L.G. et al. (1992) *Nature Gen.* 2:21-25). Thus, correction of the biologic defect in CF patients may be possible when only a small percent of the cells express CFTR. This is also consistent with our previous studies *in vitro* showing that Ad2/CFTR-1 at relatively low multiplicities of infection generated a cAMP-stimulated  $\text{Cl}^-$  secretory response in CF epithelia (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476).

This study also provides the first comprehensive data on the safety of adenovirus vectors for gene transfer to airway epithelium. Several aspects of the studies are encouraging. There was no evidence of viral replication, rather infectious viral particles were rapidly cleared from both cotton rats and primates. These data, together with our previous *in vitro* studies, suggest that replication of recombinant virus in humans will likely not be a problem. The other major consideration for safety of an adenovirus vector in the treatment of CF is the possibility of an inflammatory response. The data indicate that the virus generated an antibody response in both cotton rats and monkeys. Despite this, no evidence of a

- systemic or local inflammatory response was observed. The cells obtained by bronchoalveolar lavage and by brushing and swabs were not altered by virus application. Moreover, the histology of epithelia treated with adenovirus was indistinguishable from that of control epithelia. These data suggest that at least three sequential exposures of airway epithelium to adenovirus does not cause a detrimental inflammatory response.

These data suggest that Ad2/CFTR-1 can effectively transfer CFTR cDNA to airway epithelium and direct the expression of CFTR. They also suggest that transfer is relatively safe in animals. Thus, they suggest that Ad2/CFTR-1 may be a good vector for treating patients with CF. This was confirmed in the following example.

10

#### Example 10 - CFTR Gene Therapy in Nasal Epithelia from Human CF Subjects

### **EXPERIMENTAL PROCEDURES**

#### 15 Adenovirus vector

- The recombinant adenovirus Ad2/CFTR-1 was used to deliver CFTR cDNA. The construction and preparation of Ad2/CFTR-1, and its use *in vitro* and *in vivo* in animals, has been previously described (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476; Zabner, J. et al. (1993) *Nature Gen.* (in press)). The DNA construct comprises a full length copy of the Ad2 genome from which the early region 1 genes (nucleotides 546 to 3497) have been replaced by cDNA for CFTR. The viral E1a promoter was used for CFTR cDNA; this is a low to moderate strength promoter. Termination/polyadenylation occurs at the site normally used by E1b and protein IX transcripts. The E3 region of the virus was conserved.

#### 25 Patients

- Three patients with CF were studied. Genotype was determined by IG Labs (Framingham, MA). All three patients had mild CF as defined by an NIH score > 70 (Taussig, L.M. et al. (1973) *J. Pediatr.* 82:380-390), a normal weight for height ratio, a forced expiratory volume in one second (FEV1) greater than 50% of predicted and an arterial PO<sub>2</sub> greater than 72. All patients were seropositive for type 2 adenovirus, and had no recent viral illnesses. Pretreatment cultures of nasal swabs, pharyngeal swabs, sputum, urine, stool, and blood leukocytes were negative for adenovirus. PCR of pretreatment nasal brushings using primers for the adenovirus E1 region were negative. Patients were evaluated at least twice by FEV1, cytology of nasal mucosa, visual inspection, and measurement of V<sub>t</sub> before treatment. Prior to treatment, a coronal computed tomographic scan of the paranasal sinuses and a chest X-ray were obtained.

- The first patient was a 21 year old woman who was diagnosed at 3 months after birth. She had pancreatic insufficiency, a positive sweat chloride test (101 mEq/l), and is homozygous for the  $\Delta F508$  mutation. Her NIH score was 90 and her FEV1 was 83%



- predicted. The second patient was a 36 year old man who was diagnosed at the age of 13 when he presented with symptoms of pancreatic insufficiency. A sweat chloride test revealed a chloride concentration of 70 mEq/l. He is a heterozygote with the  $\Delta F508$  and G551D mutations. His NIH score was 88 and his FEV1 was 66% predicted. The third patient was a
- 5 50 year old woman, diagnosed at the age of 9 with a positive sweat chloride test (104 mEq/l). She has pancreatic insufficiency and insulin dependent diabetes mellitus. She is homozygous for the  $\Delta F508$  mutation. Her NIH score was 73 and her FEV1 was 65% predicted.

#### Transepithelial voltage

- 10 The transepithelial electric potential difference across the nasal epithelium was measured using techniques similar to those previously described (Alton, E.W.F.W. et al (1987) *Thorax* 42:815-817; Knowles, M. et al. (1981) *N. Eng. J. Med.* 305:1489-1495). A 23 gauge subcutaneous needle connected with sterile normal saline solution to a silver/silver chloride pellet (E.W. Wright, Guilford, CT) was used as a reference electrode. The exploring
- 15 electrode was a size 8 rubber catheter (modified Argyle<sup>R</sup> Foley catheter, St. Louis, MO) with one side hole at the tip. The catheter was filled with Ringer's solution containing (in mM), 135 NaCl, 2.4 KH<sub>2</sub>PO<sub>4</sub>, K<sub>2</sub>HPO<sub>4</sub>, 1.2CaCl<sub>2</sub>, 1.2 MgCl<sub>2</sub> and 10 Hepes (titrated to pH 7.4 with NaOH) and was connected to a silver/silver chloride pellet. Voltage was measured with a voltmeter (Keithley Instruments Inc., Cleveland, OH) connected to a strip chart recorder
- 20 (Servocorder, Watanabe Instruments, Japan). Prior to the measurements, the silver/silver chloride pellets were connected in series with the Ringer's solution; the pellets were changed if the recorded  $V_t$  was greater than  $\pm 4$  mV. The rubber catheter was introduced into the nostril under telescopic guidance (Hopkins Telescope, Karl Storz, Tuttlingen West Germany) and the side hole of the catheter was placed next to the study area in the medical aspect of the
- 25 inferior nasal turbinate. The distance from the anterior tip of the inferior turbinate and the spatial relationship with the medial turbinate, the maxillary sinus ostium, and in one patient a small polyp, were used to locate the area of Ad2/CFTR-1 administration for measurements. Photographs and video recorder images were also used. Basal  $V_t$  was recorded until no changes in  $V_t$  were observed after slow intermittent 100  $\mu$ l/min infusion of the Ringer's
- 30 solution. Once a stable baseline was achieved, 200  $\mu$ l of a Ringer's solution containing 100  $\mu$ M amiloride (Merck and Co. Inc., West Point, PA) was instilled through the catheter and changes in  $V_t$  were recorded until no further change were observed after intermittent instillations. Finally, 200  $\mu$ l Ringer's solution containing 100  $\mu$ M amiloride plus 10  $\mu$ M terbutaline (Geigy Pharmaceuticals, Ardsley, NY) was instilled and the changes in  $V_t$  were
- 35 recorded.

Measurements of basal  $V_t$  were reproducible over time: in the three treated patients, the coefficients of variation before administration of Ad2/CFTR-1 were 3.6%, 12%, and 12%. The changes induced by terbutaline were also reproducible. In 30 measurements in 9 CF patients, the terbutaline-induced changes in  $V_t$  ( $\Delta V_t$ ) ranged from 0 mV to +4 mV;

hyperpolarization of  $V_t$  was never observed. In contrast, in 7 normal subjects  $\Delta V_t$  ranged from -1 mV to -5 mV; hyperpolarization was always observed.

#### Ad2/CFTR-1 application and cell acquisition

5       The patients were taken to the operating room and monitoring was commenced using continuous EKG and pulse oximetry recording as well as automatic intermittent blood pressure measurement. After mild sedation, the nasal mucosa was anesthetized by atomizing 0.5 ml of 5% cocaine. The mucosa in the area of the inferior turbinate was then packed with cotton pledgets previously soaked in a mixture of 2 ml of 0.1% adrenaline and 8 ml of 1% tetracaine. The pledgets remained in place for 10-40 min. Using endoscopic visualization with a television monitoring system, the applicator was introduced through the nostril and positioned on the medial aspect of the inferior turbinate, at least three centimeters from its anterior tip (Figures 25A-25I). The viral suspension was infused into the applicator through connecting catheters. The position of the applicator was monitored endoscopically to ensure that it did not move and that enough pressure was applied to prevent leakage. After the virus was in contact with the nasal epithelium for thirty minutes, the viral suspension was removed, and the applicator was withdrawn. In the third patient's right nasal cavity, the virus was applied using the modified Foley catheter used for  $V_t$  measurements. The catheter was introduced without anesthetic under endoscopic guidance until the side hole of the catheter was in contact with the area of interest in the inferior turbinate. The viral solution was infused slowly until a drop of solution was seen with the telescope. The catheter was left in place for thirty minutes and then removed.

Cells were obtained from the area of virus administration approximately 2 weeks before treatment and then at weekly intervals after treatment. The inferior turbinate was packed for 10 minutes with cotton pledgets previously soaked in 1 ml of 5% cocaine. Under endoscopic control, the area of administration was gently brushed for 5 seconds. The brushed cells were dislodged in PBS. Swabs of the nasal epithelia were collected using cotton tipped applicators without anesthesia. Cytospin slides were prepared and stained with Wright's stain. Light microscopy was used to assess the respiratory epithelial cells and inflammatory cells. For biopsies, sedatives/anesthesia was administered as described for the application procedure. After endoscopic inspection, and identification of the site to be biopsied, the submucosa was injected with 1% xylocaine, with 1/100,000 epinephrine. The area of virus application on the inferior turbinate was removed. The specimen was fixed in 4% formaldehyde and stained.

#### **RESULTS**

On day one after Ad2/CFTR-1 administration and at all subsequent time points, Ad2/CFTR-1 from the nasal epithelium, pharynx, blood, urine, or stool could not be cultured. As a control for the sensitivity of the culture assay, samples were routinely spiked with 10

and 100 IU Ad2/CFTR-1. In every case, the spiked samples were positive, indicating that, at a minimum, 10 IU of Ad2/CFTR should have been detected. No evidence of a systemic response as assessed by history, physical examination, serum chemistries or cell counts, chest and sinus X-rays, pulmonary function tests, or arterial blood gases performed before and after  
5 Ad2/CFTR-1 administration. An increase in antibodies to adenovirus was not detectable by ELISA or by neutralization for 35 days after treatment.

Three to four hours after Ad2/CFTR-1 administration, at the time that local anesthesia and localized vasoconstriction abated, all patients began to complain of nasal congestion and in one case, mild rhinorrhea. These were isolated symptoms that diminished by 18 hours and  
10 resolved by 28 to 42 hours. Inspection of the nasal mucosa showed mild to moderate erythema, edema, and exudate (Figures 25A-25C). These physical findings followed a time course similar to the symptoms. The physical findings were not limited to the site of virus application, even though preliminary studies using the applicator showed that marker methylene blue was limited to the area of application. In two additional patients with CF, the  
15 identical anesthesia and application procedure were used, but saline was applied instead of virus, yet the same symptoms and physical findings were observed in these patients (Figures 25G-25I). Moreover, the local anesthesia and vasoconstriction generated similar changes even when the applicator was not used, suggesting that the anesthesia/vasoconstriction caused some, if not all the injury. Twenty-four hours after the application procedure, analysis of  
20 cells removed from nasal swabs revealed an equivalent increase in the percent neutrophils in patients treated with Ad2/CFTR-1 or with saline. One week after application, the neutrophilia had resolved in both groups. Respiratory epithelial cells obtained by nasal brushing appeared normal at one week and at subsequent time points, and showed no evidence of inclusion bodies. To further evaluate the mucosa, the epithelium was biopsied on  
25 day three in the first patient and day one in the second patient. Independent evaluation by two pathologists not otherwise associated with the study suggested changes consistent with mild trauma and possible ischemia (probably secondary to the anesthetic/vasoconstrictors used before virus administration), but there were no abnormalities suggestive of virus-mediated damage.

30 Because the application procedure produced some mild injury in the first two patients, the method of administration was altered in the third patient. The method used did not require the use of local anesthesia or vasoconstriction and which was thus less likely to cause injury, but which was also less certain in its ability to constrain Ad2/CFTR-1 in a precisely defined area. On the right side, Ad2/CFTR-1 was administered as in the first two patients,  
35 and on the left side, the virus was administered without anesthesia or the applicator, instead using a small Foley catheter to apply and maintain Ad2/CFTR-1 in a relatively defined area by surface tension (Figure 25E). On the right side, the symptoms and physical findings were the same as those observed in the first two patients. By contrast, on the left side there were no symptoms and on inspection the nasal mucosa appeared normal (Figures 25D-25F). Nasal

swabs obtained from the right side showed neutrophilia similar to that observed in the first two patients. In contrast, the left side which had no anesthesia and minimal manipulation, did not develop neutrophilia. Biopsy of the left side on day 3 after administration (Figure 26), showed morphology consistent with CF-- a thickened basement membrane and occasional polymorphonuclear cells in the submucosa-- but no abnormalities that could be attributed to the adenovirus vector.

The first patient developed symptoms of a sore throat and increased cough that began three weeks after treatment and persisted for two days. Six weeks after treatment she developed an exacerbation of her bronchitis/bronchiectasis and hemoptysis that required hospitalization. The second patient had a transient episode of minimal hemoptysis three weeks after treatment; it was not accompanied by any other symptoms before or after the episode. The third patient has an exacerbation of bronchitis three weeks after treatment for which she was given oral antibiotics. Based on each patient's pretreatment clinical history, evaluation of the episodes, and viral cultures, no evidence could be discerned that linked these episodes to administration of Ad2/CFTR-1. Rather the episodes appeared consistent with the normal course of disease in each individual.

The loss of CFTR  $\text{Cl}^-$  channel function causes abnormal ion transport across affected epithelia, which in turn contributes to the pathogenesis of CF-associated airway disease (Boat, T.F. et al. in *The Metabolic Basis of Inherited Diseases* (Scriver, C.R. et al. eds., McGraw-Hill, New York (1989); Quinton, P.M. (1990) *FASEB J.* 4:2709-2717). In airway epithelia, ion transport is dominated by two electrically conductive processes: amiloride-sensitive absorption of  $\text{Na}^+$  from the mucosal to the submucosal surface and cAMP-stimulated  $\text{Cl}^-$  secretion in the opposite direction. (Quinton, P.M. (1990) *FASEB J.* 4:2709-2717; Welsh, M.J. (1987) *Physiol. Rev.* 67:1143-1184). These two transport processes can be assessed noninvasively by measuring the voltage across the nasal epithelium ( $V_t$ ) *in vivo* (Knowles, M. et al (1981) *N. Eng. J. Med.* 305:1489-1495; Alton, E.W.F.W. et al.(1987) *Thorax* 42:815-817). Figure 27 shows an example from a normal subject. Under basal conditions,  $V_t$  was electrically negative (lumen referenced to the submucosal surface). Perfusion of amiloride (100  $\mu\text{M}$ ) onto the mucosal surface inhibited  $V_t$  by blocking apical  $\text{Na}^+$  channels (Knowles, M. et al (1981) *N. Eng. J. Med.* 305:1489-1495; Quinton, P.M. (1990) *FASEB J.* 4:2709-2717; Welsh, M.J. (1992) *Neuron* 8:821-829). Subsequent perfusion of terbutaline (10  $\mu\text{M}$ ) a  $\beta$ -adrenergic agonist, hyperpolarized  $V_t$  by increasing cellular levels of cAMP, opening CFTR  $\text{Cl}^-$  channels, and stimulating chloride secretion (Quinton, P.M. (1990) *FASEB J.* 4:2709-2717; Welsh, M.J. et al. (1992) *Neuron* 8:821-829). Figure 28A shows results from seven normal subjects: basal  $V_t$  was  $-10.5 \pm 1.0\text{mV}$ , and in the presence of amiloride, terbutaline hyperpolarized  $V_t$  by  $-2.3 \pm 0.5\text{mV}$ .

In patients with CF,  $V_t$  was more electrically negative than in normal subjects (Figure 28B), as has been previously reported (Knowles, M. et al. (1981) *N. Eng. J. Med.* 305:1489-1495). Basal  $V_t$  was  $-37.0 \pm 2.4\text{mV}$ , much more negative than values in normal subjects ( $P <$

0.001). (Note the difference in scale in Figure 28A and Figure 28B). Amiloride inhibited  $V_t$ , as it did in normal subjects. However,  $V_t$  failed to hyperpolarize when terbutaline was perfused onto the epithelium in the presence of amiloride. Instead,  $V_t$  either did not change or became less negative: on average  $V_t$  depolarized by  $+1.8 \pm 0.6$  mV, a result very different from that observed in normal subjects. ( $P < 0.001$ ).

After Ad2/CFTR-1 was applied, basal  $V_t$  became less negative in all three CF patients: Figure 29A shows an example from the third patient before (Figure 29A) and after (Figure 29B) treatment and Figures 30A, 30C, and 30E show the time course of changes in basal  $V_t$  for all three patients. The decrease in basal  $V_t$  suggests that application of Ad2/CFTR-1 corrected the CF electrolyte transport defect in nasal epithelium of all three patients. Additional evidence came from an examination of the response to terbutaline. Figure 30B shows that in contrast to the response before Ad2/CFTR-1 was applied, after virus replication, in the presence of amiloride, terbutaline stimulated  $V_t$ . Figures 30B, 30D, and 30F show the time course of the response. These data indicate that Ad2/CFTR-1 corrected the CF defect in  $Cl^-$  transport. Correction of the  $Cl^-$  transport defect cannot be attributed to the anesthesia/application procedure because it did not occur in patients treated with saline instead of Ad2/CFTR-1 (Figure 31). Moreover, the effects of the anesthesia were generalized on the nasal mucosa, but basal  $V_t$  decreased only in the area of virus administration. Finally, similar changes were observed in the left nasal mucosa of the third patient (Figures 30E and 30F), which had no symptomatic or physical response after the modified application procedure.

Unsuccessful attempts were made to detect CFTR transcripts by reverse transcriptase-PCR and by immunocytochemistry in cells from nasal brushings and biopsies. Although similar studies in animals have been successful (Zabner, J. et al. (1993) *Nature Gen.* (in press)), those studies used much higher doses of Ad2/CFTR-1. The lack of success in the present case likely reflects the small amount of available tissue, the low MOI, the fact that only a fraction of cells may have been corrected, and the fact that Ad2/CFTR-1 contains a low to moderate strength promoter (Ela) which produces much less mRNA and protein than comparable constructs using a much stronger CMV promoter (unpublished observation). The Ela promoter was chosen because CFTR normally expressed at very low levels in airway epithelial cells (Trapnell, B.C. et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:6565-6569). It is also difficult to detect CFTR protein and mRNA in normal human airway epithelia, although function is readily detected because a single ion channel can conduct a very large number of ions per second and thus efficiently support  $Cl^-$  transport.

With time, the electrical changes that indicate correction of the CF defect reverted toward pretreatment values. However, the basal  $V_t$  appeared to revert more slowly than did the change in  $V_t$  produced by terbutaline. The significance of this difference is unknown, but it may reflect the relative sensitivity of the two measurements to expression of normal CFTR. In any case, this study was not designed to test the duration of correction because the treated

area was removed by biopsy on one side and the nasal mucosa on the other side was brushed to obtain cells for analysis at 7 to 10 days after virus administration, and then at approximately weekly intervals. Brushing the mucosa removes cells, disrupts the epithelium, and reduces basal  $V_t$  to zero for at least two days afterwards, thus preventing an accurate  
5 assessment of duration of the effect of Ad2/CFTR-1.

Efficacy of adenovirus-mediated gene transfer.

The major conclusion of this study is that *in vivo* application of a recombinant adenovirus encoding CFTR can correct the defect in airway epithelial  $Cl^-$  transport that is  
10 characteristic of CF epithelia.

Complementation of the  $Cl^-$  channel defect in human nasal epithelium could be measured as a change in basal voltage and as a change in the response to cAMP agonists. Although the protocol was not designed to establish duration, changes in these parameters were detected for at least three weeks. These results represent the first report that  
15 administration of a recombinant adenovirus to humans can correct a genetic lesion as measured by a functional assay. This study contrasts with most earlier attempts at gene transfer to humans, in that a recombinant viral vector was administered directly to humans, rather than using a *in vitro* protocol involving removal of cells from the patient, transduction of the cells in culture, followed by reintroduction of the cells into the patient.

20 Evidence that the CF  $Cl^-$  transport defect was corrected at all three doses of virus, corresponding to 1, 3, and 25 MOI, was obtained. This result is consistent with earlier studies showing that similar MOIs reversed the CF fluid and electrolyte transport defects in primary cultures of CF airway cells grown as epithelia on permeable filter supports (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476 and Zabner et al. submitted for  
25 publication): at an MOI of less than 1, cAMP-stimulated  $Cl^-$  secretion was partially restored, and after treatment with 1 MOI Ad2/CFTR-1 cAMP agonists stimulated fluid secretion that was within the range observed in epithelia from normal subjects. At an MOI of 1, a related adenovirus vector produced  $\beta$ -galactosidase activity in 20% of infected epithelial cells as assessed by fluorescence-activated cell analysis (Zabner et al. submitted for publication).  
30 Such data would imply that pharmacologic dose of adenovirus in CF airways might correspond to an MOI of one. If it is estimated that there are  $2 \times 10^6$  cells/cm<sup>2</sup> in the airway (Mariassy, A.T. in *Comparative Biology of the Normal Lung* (CRC Press, Boca Raton 1992), and that the airways from the trachea to the respiratory bronchioles have a surface area of 1400 cm<sup>2</sup> (Weibel, E.R. *Morphometry of the Human Lung* (Springer Verlag, Heidelberg,  
35 1963) then there would be approximately  $3 \times 10^9$  potential target cells. Assuming a particle to IU ratio of 100, this would correspond to approximately  $3 \times 10^{11}$  particles of adenovirus with a mass of approximately 75  $\mu$ g. While obviously only a crude estimate, such information is useful in designing animal experiments to establish the likely safety profile of a human dose.

It is possible that an efficacious MOI of recombinant adenovirus could be less than the lowest MOI tested here. Some evidence suggests that not all cells in an epithelial monolayer need to express CFTR to correct the CF electrolyte transport defects. Mixing experiments showed that when perhaps 5-10% of cells overexpress CFTR, the monolayer exhibits wild-type electrical properties (Johnson, L.G. et al. (1992) *Nature Gen.* 2:21-25). Studies using liposomes to express CFTR in mice bearing a disrupted CFTR gene also suggest that only a small proportion of cells need to be corrected (Hyde, S.C. et al. (1993) *Nature* 362:250-255). The results referred to above using airway epithelial monolayers and multiplicities of Ad2/CFTR-1 as low as 0.1 showed measurable changes in  $\text{Cl}^-$  secretion (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476 and Zabner et al. submitted for publication).

Given the very high sensitivity of electrolyte transport assays (which result because a single  $\text{Cl}^-$  channel is capable of transporting large numbers of ions/sec) and the low activity of the E1a promoter used to transcribe CFTR, the inability to detect CFTR protein and CFTR mRNA are perhaps not surprising. Although CFTR mRNA could not be detected by reverse transcriptase-PCR, Ad2/CFTR-1 DNA could be detected in the samples by standard PCR, demonstrating the presence of input DNA and suggesting that the reverse transcriptase reaction may have been suboptimal. This could have occurred because of factors in the tissue that inhibit the reverse transcriptase. Although there is little doubt that the changes in electrolyte transport measured here result from expression of CFTR, it remains to be seen whether this will lead to measurable clinical changes in lung function.

#### Safety considerations.

Application of the adenovirus vector to the nasal epithelium in these three patients was well-tolerated. Although mild inflammation was observed in the nasal epithelium of all three patients following administration of Ad2/CFTR-1, similar changes were observed in two volunteers who underwent a sham procedure using saline rather than the viral vector. Clearly a combination of anesthetic- and procedure-related trauma resulted in the changes in the nasal mucosa. There is insufficient evidence to conclude that no inflammation results from virus administration. However, using a modified administration of the highest MOI of virus tested (25 MOI) in one patient, no inflammation was observed under conditions that resulted in evidence of biophysical efficacy that lasted until the area was removed by biopsy at three days.

There was no evidence of replication of Ad2/CFTR-1. Earlier studies had established that replication of Ad2/CFTR-1 in tissue culture and experimental animals is severely impaired (Rich, D.P. et al. (1993) *Human Gene Therapy* 4:461-476; Zabner, J. et al. (1993) *Nature Gen.* (in press)). Replication only occurs in cells that supply the missing early proteins of the E1 region of adenovirus, such as 293 cells, or under conditions where the E1 region is provided by coinfection with or recombination with an E1-containing adenovirus

(Graham, F.L. and Prevec, L. Vaccines: New Approaches to Immunological Problems (R.W. Ellis, ed., Boston, Butterworth-Heinemann, 1992); Berkner, K.L. (1988) *Biotechniques* 6:616-629). The patients studied here were seropositive for adenovirus types 2 and 5 prior to the study were negative for adenovirus upon culture of nasal swabs prior to administration of Ad2/CFTR-1, and were shown by PCR methods to lack endogenous E1 DNA sequences such as have been reported in some human subjects (Matsuse T. et al. (1992) *Am. Rev. Respir. Dis.* 146:177-184).

#### Example 11 - Construction and Packaging of Pseudo Adenoviral Vector (PAV)

10 With reference to Figure 32, the PAV construct was made by inserting the Ad2 packaging signal and E1 enhancer region (0-358 nt) in Bluescript II SK- (Stratagene, LaJolla, CA). A variation of this vector, known as PAV II was constructed similarly, except the Ad2 packaging signal and E1 enhancer region contained 0-380 nt. The addition of nucleotides at the 5' end results in larger PAVs, which may be more efficiently packaged, yet would include  
15 more adenoviral sequences and therefore could potentially be more immunogenic or more capable of replicating.

To allow ease of manipulation for either the insertion of gene coding regions or complete excision and use in transfections for the purpose of generating infectious particles, a complementary plasmid was also built in pBluescript SKII-. This complementary plasmid  
20 contains the Ad2 major late promoter (MLP) and tripartite leader (TPL) DNA and an SV40 T-antigen nuclear localization signal (NLS) and polyadenylation signal (SVpA). As can be seen in Figure 32, this plasmid contains a convenient restriction site for the insertion of genes of interest between the MLP/TPL and SV40 poly A. This construct is engineered such that the entire cassette may be excised and inserted into the former PAV I or PAV II construct.

25 Generation of PAV infectious particles was performed by excision of PAV from the plasmid with the Apa I and Sac II restriction endonucleases and co-transfection into 293 cells (an Ela/Elb expressing cell line) (Graham, F.L. et al, (1977) *J. Gen Virol* 36:59-74) with either wild-type Ad2, or packaging/replication deficient helper virus. Purification of PAV from helper can be accompanied by CsCl gradient isolation as PAV viral particles will be of a  
30 lower density and will band at a higher position in the gradient.

For gene therapy, it is desirable to generate significant quantities of PAV virion free from contaminating helper virus. The primary advantage of PAV over standard adenoviral vectors is the ability to package large DNA inserts into virion (up to about 36 kb). However, PAV requires a helper virus for replication and packaging and this helper virus will be the  
35 predominant species in any PAV preparation. To increase the proportion of PAV in viral preparation several approaches can be employed. For example, one can use a helper virus which is partially defective for packaging into virions (either by virtue of mutations in the packaging sequences (Grable, M. and Hearing P. (1992) *J. Virol.* 66: 723-731)) or by virtue of its size -viruses with genome sizes greater than approximately 37.5 kb package



inefficiently. In mixed infections with packaging defective virus, PAV would be expected to be represented at higher levels in the virus mixture than would occur with non-packaging defective helper viruses.

Another approach is to make the helper virus dependent upon PAV for its own replication. This may most easily be accomplished by deleting an essential gene from the helper virus (e.g. IX or a terminal protein) and placing that gene in the PAV vector. In this way neither PAV nor the helper virus is capable of independent replication - PAV and the helper virus are therefore co-dependent. This should result in higher PAV representation in the resulting virus preparation.

A third approach is to develop a novel packaging cell line, which is capable of generating significant quantities of PAV virion free from contaminating helper virus. A novel protein IX, (pIX) packaging system has been developed. This system exploits several documented features of adenovirus molecular biology. The first is that adenoviral defective particles are known to comprise up to 30% or more of standard wild-type adenoviral preparations. These defective or incomplete particles are stable and contain 15-95% of the adenoviral genome, typically 15-30%. Packaging of a PAV genome (15-30% of wild-type genome) should package comparably. Secondly, stable packaging of full-length Ad genome but not genomes <95% required the presence of the adenoviral gene designated pIX.

The novel packaging system is based on the generation of an Ad protein pIX expressing 293 cell line. In addition, an adenoviral helper virus engineered such that the E1 region is deleted but enough exogenous material is inserted to equal or slightly exceed the full length 36 kb size. Both of these two constructs would be introduced into the 293/pIX cell line as purified DNA. In the presence of pIX, yields of both predicted progeny viruses as seen in current PAV/Ad2 production experiments can be obtained. Virus containing lysates from these cells can then be titered independently (for the marker gene activity specific to either vector) and used to infect standard 293 (lacking pIX) at a multiplicity of infection of 1 relative to PAV. Since research with this line as well as from incomplete or defective particle research indicates that full length genomes have a competitive packaging advantage, it is expected that infection with an MOI of 1 relative to PAV will necessarily equate to an effective MOI for helper of greater than 1. All cells will presumably contain both PAV (at least 1) and helper (greater than 1). Replication and viral capsid production in this cell should occur normally but only PAV genomes should be packaged. Harvesting these 293/pIX cultures is expected to yield essentially helper-free PAV.

#### Example 12 - Construction of Ad2-E4/ORF 6

Ad2-E4/ORF6 (Figure 33 shows the plasmid construction of Ad2-E4/ORF6) which is an adenovirus 2 based vector deleted for all Ad2 sequences between nucleotides 32815 and 35577. This deletion removes all open reading frames of E4 but leaves the E4 promoter and first 32-37 nucleotides of the E4 mRNA intact. In place of the deleted sequences, a DNA

fragment encoding ORF6 (Ad2 nucleotides 34082-33178) which was derived by polymerase chain reaction of Ad2 DNA with ORF6 specific DNA primers (Genzyme oligo. # 2371 - CGGATCCTTTATTATAGGGGAAGTCCACGCCTAC (SEQ. ID NO:8) and oligo. #2372 - CGGGATCCATCGATGAAATATGACTACGTCCG (SEQ. ID NO:9) were inserted). Additional sequences supplied by the oligonucleotides included a cloning site at the 5' and 3' ends of the PCR fragment (Clal and BamHI respectively) and a polyadenylation sequence at the 3' end to ensure correct polyadenylation of the ORF6 mRNA. As illustrated in Figure 33, the PCR fragment was first ligated to a DNA fragment including the inverted terminal repeat (ITR) and E4 promoter region of Ad2 (Ad2 nucleotides 35937-35577) and cloned in the bacterial plasmid pBluescript (Stratagene) to create plasmid ORF6. After sequencing to verify the integrity of the ORF6 reading frame, the fragment encompassing the ITR and ORF6 was subcloned into a second plasmid, pAd Δ E4, which contains the 3' end of Ad2 from a Sac I site to the 3' ITR (Ad2 nucleotides 28562-35937) and is deleted for all E4 sequences (promoter to poly A site Ad2 positions 32815-35641) using flanking restriction sites. In this second plasmid, virus expressing only E4 ORF6, pAdORF6 was cut with restriction enzyme PacI and ligated to Ad2 DNA digested with PacI. This PacI site corresponds to Ad2 nucleotide 28612. 293 cells were transfected with the ligation and the resulting virus was subjected to restriction analysis to verify that the Ad2 E4 region had been substituted with the corresponding region of pAdORF6 and that the only remaining E4 open reading frame was ORF6.

A cell line could in theory be established that would fully complement E4 functions deleted from a recombinant virus. The problem with this approach is that E4 functions in the regulation of host cell protein synthesis and is therefore toxic to cells. The present recombinant adenoviruses are deleted for the E1 region and must be grown in 293 cells which complement E1 functions. The E4 promoter is activated by the Ela gene product, and therefore to prevent inadvertent toxic expression of E4 transcription of E4 must be tightly regulated. The requirements of such a promoter or transactivating system is that in the uninduced state expression must be low enough to avoid toxicity to the host cell, but in the induced state must be sufficiently activated to make enough E4 gene product to complement the E4 deleted virus during virus production.

#### Example 13

An adenoviral vector is prepared as described in Example 7 while substituting the phosphoglycerate kinase (PGK) promoter for the Ela promoter.

#### Example 14

An adenoviral vector is prepared as described in Example 11 while substituting the PGK promoter for the Ad2 major late promoter (MLP).

Example 15: Generation of Ad2-ORF6/PGK-CFTR

This protocol uses a second generation adenovirus vector named Ad2-ORF6/PGK-CFTR. This virus lacks E1 and in its place contains a modified transcription unit with the PGK promoter and a poly A addition site flanking the CFTR cDNA. The PGK promoter is of only moderate strength but is long lasting and not subject to shut off. The E4 region of the vector has also been modified in that the whole coding sequence has been removed and replaced by ORF6, the only E4 gene essential for growth of Ad in tissue culture. This has the effect of generating a genome of 101% the size of wild type Ad2.

The DNA construct comprises a full length copy of the Ad2 genome from which the early region 1 (E1) genes (present at the 5' end of the viral genome) have been deleted and replaced by an expression cassette encoding CFTR. The expression cassette includes the promoter for phosphoglycerate kinase (PGK) and a polyadenylation (poly A) addition signal from the bovine growth hormone gene (BGH). In addition, the E4 region of Ad2 has been deleted and replaced with only open reading frame 6 (ORF6) of the Ad2 E4 region. The adenovirus vector is referred to as AD2-ORF6/PGK-CFTR and is illustrated schematically in Figure 34. The entire wild-type Ad2 genome has been previously sequenced (Roberts, R.J., (1986) In Adenovirus DNA, W. Oberfler, editor, Martinus Nihoff Publishing, Boston) and the existing numbering system has been adopted here when referring to the wild type genome. Ad2 genomic regions flanking E1 and E4 deletions, and insertions into the genome are being completely sequenced.

The Ad2-ORF6/PGK-CFTR construct differs from the one used in our earlier protocol (Ad2/CFTR-1) in that the latter utilized the endogenous E1a promoter, had no poly A addition signal directly downstream of CFTR and retained an intact E4 region. The properties of Ad2/CFTR-1 in tissue culture and in animal studies have been reported (Rich et al., (1993) *Human Gene Therapy* 4:461-467; and Zabner et al. (1993) *Nature Genetics* (in Press)).

At the 5' end of the genome, nucleotides 357 to 3328 of Ad2 have been deleted and replaced with (in order 5' to 3') 22 nucleotides of linker, 534 nucleotides of the PGK promoter, 86 nucleotides of linker, nucleotides 123-4622 of the published CFTR sequence (Riordan et al. (1989) *Science* 245:1066-1073), 21 nucleotides of linker, and a 32 nucleotide synthetic BGH poly A addition signal followed by a final 11 nucleotides of linker. The topology of the 5' end of the recombinant molecule is illustrated in Figure 34.

At the 3' end of the genome of Ad2-ORF6/PGK-CFTR, Ad2 sequences between nucleotides 32815 and 35577 have been deleted to remove all open reading frames of E4 but retain the E4 promoter, the E4 cap sites and first 32-37 nucleotides of E4 mRNA. The deleted sequences were replaced with a fragment derived by PCR which contains open reading frame 6 of Ad2 (nucleotides 34082-33178) and a synthetic poly A addition signal. The topology of the 3' end of the molecule is shown in Figure 34. The sequence of this segment of the molecule will be confirmed. The remainder of the Ad2 viral DNA sequence is

published in Roberts, R.J. in Adenovirus DNA. (W. Oberfler, Martinus Nihoff Publishing, Boston, 1986 ). The overall size of the Ad2-ORF6/PGK-CFTR vector is 36,336 bp which is 101.3% of full length Ad2. See Table III for the sequence of Ad2-ORF6/PGK-CFTR.

The CFTR transcript is predicted to initiate at one of three closely spaced  
5 transcriptional start sites in the cloned PGK promoter (Singer-Sam et al. (1984) *Gene* 32:409-417) at nucleotides 828, 829 and 837 of the recombinant vector (Singer-Sam et al. (1984) *Gene* 32:409-417). A hybrid 5' untranslated region is comprised of 72, 80 or 81 nucleotides of PGK promoter region, 86 nucleotide of linker sequence, and 10 nucleotides derived from the CFTR insert. Transcriptional termination is expected to be directed by the BGH poly A  
10 addition signal at recombinant vector nucleotide 5530 yielding an approximately 4.7 kb transcript. The CFTR coding region comprises nucleotides 1010-5454 of the recombinant virus and nucleotides 182, 181 or 173 to 4624, 4623, or 4615 of the PGK-CFTR-BGH mRNA respectively, depending on which transcriptional initiation site is used. Within the CFTR cDNA there are two differences from the published (Riordan et al, *cited supra*) cDNA  
15 sequence. An A to C change at position 1990 of the CFTR cDNA (published CFTR cDNA coordinates) which was an error in the original published sequence, and a T to C change introduced at position 936. The change at position 936 is translationally silent but increases the stability of the cDNA when propagated in bacterial plasmids (Gregory et al. (1990) *Nature* 347:382-386; and Cheng et al. (1990) *Cell* 63:827-834). The 3' untranslated region of  
20 the predicted CFTR transcript comprises 21 nucleotides of linker sequence and approximately 10 nucleotides of synthetic BGH poly A additional signal.

Although the activity of CFTR can be measured by electrophysiological methods, it is relatively difficult to detect biochemically or immunocytochemically, particularly at low levels of expression (Gregory et al., *cited supra*; and Denning et al. (1992) *J. Cell Biol.*  
25 118:551-559). A high expression level reporter gene encoding the *E. coli*  $\beta$  galactosidase protein fused to a nuclear localization signal derived from the SV40 T-antigen was therefore constructed. Reporter gene transcription is driven by the powerful CMV early gene constitutive promoter. Specifically, the E1 region of wild type Ad2 between nucleotides 357-3498 has been deleted and replaced it with a 515 bp fragment containing the CMV promoter  
30 and a 3252 bp fragment encoding the  $\beta$  galactosidase gene.

#### Regulatory Characteristics of the Elements of the AD2-ORF6/PGK-CFTR

In general terms, the vector is similar to several earlier adenovirus vectors encoding CFTR but it differs in three specific ways from the Ad2/CFTR-1 construct.

35

#### PGK Promoter

Transcription of CFTR is from the PGK promoter. This is a promoter of only moderate strength but because it is a so-called house keeping promoter we considered it more likely to be capable of long term albeit perhaps low level expression. It may also be less

likely to be subject to "shut-down" than some of the very strong promoters used in other studies especially with retroviruses. Since CFTR is not an abundant protein longevity of expression is probably more critical than high level expression. Expression from the PGK promoter in a retrovirus vector has been shown to be long lasting (Apperley et al. (1991) *Blood* 78:310-317).

#### Polyadenylation Signal

Ad2-ORG6/PGK-CFTR contains an exogenous poly A addition signal after the CFTR coding region and prior to the protein IX coding sequence of the Ad2 E1 region. Since protein IX is believed to be involved in packaging of virions, this coding region was retained. Furthermore, since protein IX is synthesized from a separate transcript with its own promoter, to prevent possible promoter occlusion at the protein IX promoter, the BGH poly A addition signal was inserted. There is indirect evidence that promoter occlusion can be problematic in that Ad2/CMV  $\beta$ Gal grows to lower viral titers on 293 cells than does Ad2/ $\beta$ gal-1. These constructs are identical except for the promoter used for  $\beta$  galactosidase expression. Since the CMV promoter is much stronger than the E1a promoter it is probable that abundant transcription from the CMV promoter through the  $\beta$  galactosidase DNA into the protein IX coding region reduces expression of protein IX from its own promoter by promoter occlusion and that this is responsible for the lower titer of Ad2/CMV- $\beta$ gal obtained.

#### Alterations of the E4 Region

A large portion of the E4 region of the Ad2 genome has been deleted for two reasons. The first reason is to decrease the size of the vector used or expression of CFTR. Adenovirus vectors with genomes much larger than wild type are packaged less efficiently and are therefore difficult to grow to high titer. The combination of the deletions in the E1 and E4 regions in Ad2-ORG6/PGK-CFTR reduce the genome size to 101% of wild type. In practice it is straightforward to prepare high titer lots of this virus.

The second reason to remove E4 sequences relates to the safety of adenovirus vectors. A goal of these studies is to remove as many viral genes as possible to inactivate the Ad2 virus backbone in as many ways as possible. The OF 6/7 gene of the E4 region encodes a protein that is involved in activation of the cellular transcription factor E2-F which is in turn implicated in the activation of the E2 region of adenovirus (Hemstrom et al. (1991) *J. Virol.* 65:1440-1449). Therefore removal of ORF6/7 from adenovirus vectors may provide a further margin of safety at least when grown in non-proliferating cells. The removal of the E1 region already renders such vectors disabled, in part because E1a, if present, is able to displace E2-F from the retinoblastoma gene product, thereby also contributing to the stimulation of E2 transcription. The ORF6 reading frame of Ad2 was added back to the E1-E4 backbone of the Ad2-ORG6/PGK-CFTR vector because ORF6 function is essential for production of the recombinant virus in 293 cells. ORF6 is believed to be involved in DNA replication, host

cell shut off and late mRNA accumulation in the normal adenovirus life cycle. The E1-E4-ORF6<sup>+</sup> backbone Ad2 vector does replicate in 293 cells.

The promoter/enhancer use to drive transcription of ORF6 of E4 is the endogenous E4 promoter. This promoter requires E1a for activation and contains E1a core enhancer elements and SP1 transcription factor binding sites (reviewed in Berk, A.J. (1986) *Ann. Rev. Genet.* 20:75-79).

#### Replication Origin

The only replication origins present in Ad2-ORF6/PGK-CFTR are those present in the Ad2 parent genome. Replication of Ad2-ORF6/PGK-CFTR sequences has not been detected except when complemented with wild type E1 activity.

#### Steps Used to Derive the DNA Construct

Construction of the recombinant Ad2-ORF6/PGK-CFTR virus was accomplished by *in vivo* recombination of Ad2-ORF6 DNA and a plasmid containing the 5' 10.7 kb of adenovirus engineered to have an expression cassette encoding the human CFTR cDNA driven by the PGK promoter and a BGH poly A signal in place of the E1 coding region.

The generation of the plasmid, pBRAd2/PGK-CFTR is described here. The starting plasmid contains an approximately 7.5 kb insert cloned into the ClaI and BamHI sites of pBR322 and comprises the first 10,680 nucleotides of Ad2 with a deletion of the Ad2 sequences between nucleotides 356 and 3328. This plasmid contains a CMV promoter inserted into the ClaI and SpeI sites at the region of the E1 deletion and is designated pBRAd2/CMV. The plasmid also contains the Ad2 5' ITR, packaging and replication sequences and E1 enhancer. The E1 promoter, E1a and most of E1b coding region has been deleted. The 3' terminal portion of the E1b coding region coincides with the pIX promoter which was retained. The CMV promoter was removed and replaced with the PGK promoter as a ClaI and SpeI fragment from the plasmid PGK-GCR. The resulting plasmid, pBRAd2/PGK, was digested with AvrII and BstBI and the excised fragment replaced with the SpeI to BstBI fragment from the plasmid construct pAd2E1a/CFTR. This transferred a fragment containing the CFTR cDNA, BGH poly A signal and the Ad2 genomic sequences from 3327 to 10,670. The resulting plasmid is designated pBRAd2/PGK-CFTR. The CFTR cDNA fragment was originally derived from the plasmid pCMV-CFTR-936C using restriction enzymes SpeI and Ecl136II. pCMV-CFTR-936C consists of a minimal CFTR cDNA encompassing nucleotides 123-4622 of the published CFTR sequence cloned into the multiple cloning site of pRC/CMV (Invitrogen Corp.) using synthetic linkers. The CFTR cDNA within this plasmid has been completely sequenced.

The Ad2 backbone virus with the E4 region that expresses only open reading frame 6 was constructed as follows. A DNA fragment encoding ORF6 (Ad2 nucleotides 34082-33178) was derived by PCR with ORF6 specific DNA primers. Additional sequences

- 60 -

supplied by the oligonucleotides include cloning sites at the 5' and 3' ends of the PCR fragment. (ClaI and BamHI respectively) and a poly A addition sequence AATAAA at the 3' end to ensure correct polyadenylation of ORF6 mRNA. The PCR fragment was cloned into pBluescript (Stratagene) along with an Ad2 fragment (nucleotides 35937-35577) containing the inverted terminal repeat, E4 promoter, E4 mRNA cap sites and first 32-37 nucleotides of E4 mRNA to create pORF6. A Sall-BamHI fragment encompassing the ITR and ORF6 was used to replace the Sall-BamHI fragment encompassing the ITR and E4 deletion in pAdΔE4 contains the 3' end of Ad2 from a SpeI site to the 3' ITR (nucleotides 27123-35937) and is deleted for all E4 sequences including the promoter and poly A signal (nucleotides 32815-35641). The resulting construct, pAdE4ORF6 was cut with PacI and ligated to Ad2 DNA digested with PacI nucleotide 28612). 293 cells were transfected with the ligation reaction to generate virus containing only open reading frame 6 from the E4 region.

#### In Vitro Studies with Ad2-ORF6/PGK-CFTR

The ability of Ad2-ORF6/PGK-CFTR to express CFTR in several cell lines, including human HeLa cells, human 293 cells, and primary cultures of normal and CF human airway epithelia was tested. As an example, the results from the human 293 cells is related here. When human 293 cells were grown on culture dishes, the vector was able to transfer CFTR cDNA and express CFTR as assessed by immunoprecipitation and by functional assays of halide efflux. Gregory, R.J. et al. (1990) *Nature* 347:382-386; Cheng, S.H. et al. (1990) *Cell* 63:827-834. More specifically, procedures for preparing cell lysates, immunoprecipitation of proteins using anti-CFTR antibodies, one-dimensional peptide analysis and SDS-polyacrylamide gel electrophoresis were as described by Cheng et al. Cheng, S.H. et al. (1990) *Cell* 63:827-834. Halide efflux assays were performed as described by Cheng, S.H. et al. (1991) *Cell* 66:1027-1036. cAMP-stimulated CFTR chloride channel activity was measured using the halide sensitive fluorophore SPQ in 293 cells treated with 500 IU/cell Ad2-ORF6/PGK-CFTR. Stimulation of the infected cells with forskolin (20 μM) and IBMX (100 μM) increased SPQ fluorescence indicating the presence of functional chloride channels produced by the vector.

Additional studies using primary cultures of human airway (nasal polyp) epithelial cells (from CF patients) infected with Ad2-ORF6/PGK-CFTR demonstrated that Ad2-ORF6/PGK-CFTR infection of the nasal polyp epithelial cells resulted in the expression of cAMP dependent Cl<sup>-</sup> channels. Figure 35 is an example of the results obtained from such studies. Primary cultures of CF nasal polyp epithelial cells were infected with Ad2-ORF6/PGK-CFTR at multiplicities of 0.3, 3, and 50. Three days post infection, monolayers were mounted in Ussing chambers and short-circuit current was measured. At the indicated times: (1) 10 μM amiloride, (2) cAMP agonists (10 μM forskolin and 100 μM IBMX), and (3) 1 mM diphenylamine-2-carboxylate were added to the mucosal solution.

In Vivo Studies with Ad2-ORF6/PGK-CFTRVirus preparation

- Two preparations of Ad2-ORF6/PGK-CFTR virus were used in this study. Both were prepared at Genzyme Corporation, in a Research Laboratory. The preparations were purified on a CsCl gradient and then dialyzed against tris-buffered saline to remove the CsCl. The preparation for the first administration (lot #2) had a titer of  $2 \times 10^{10}$  IU/ml. The preparation for the second administration (lot #6) had a titer of  $4 \times 10^{10}$  IU/ml.

10 Animals

- Three female Rhesus monkeys, *Macaca mulatta*, were used for this study. Monkey C (#20046) weighed 6.4 kg. Monkey D (#20047) weighed 6.25 kg. Monkey E (#20048) weighed 10 kg. The monkeys were housed in the University of Iowa at least 360 days before the start of the study. The animals were maintained with free access to food and water throughout the study. The animals were part of a safety study and efficacy study for a different viral vector (Ad2/CFTR-1) and they were exposed to 3 nasal viral instillation throughout the year. The previous instillation of Ad2/CFTR-1 was performed 116 days prior to the initiation of this study. All three Rhesus monkeys had an anti-adenoviral antibody response as detected by ELISA after each viral instillation. There are no known contaminants that are expected to interfere with the outcome of this study. Fluorescent lighting was controlled to automatically provide alternate light/dark cycles of approximately 12 hours each. The monkeys were housed in an isolation room in separate cages. Strict respiratory and body fluid isolation precautions were taken.

25 Virus administration

- For application of the virus, the monkeys were anesthetized by intramuscular injection of ketamine (15 mg/kg). The entire epithelium of one nasal cavity in each monkey was used for this study. A foley catheter (size 10) was inserted through each nasal cavity into the pharynx, the balloon was inflated with a 2-3 ml of air, and then pulled anteriorly to obtain a tight occlusion at the posterior choana. The Ad2-ORF6/PGK-CFTR virus was then instilled slowly into the right nostril with the posterior balloon inflated. The viral solution remained in contact with the nasal mucosa for 30 min. The balloons were deflated, the catheters were removed, and the monkeys were allowed to recover from anesthesia.

- On the first administration, the viral preparation had a titer of  $2 \times 10^{10}$  IU/ml and each monkey received approximately 0.3 ml. Thus the total dose applied to each monkey was approximately  $6.5 \times 10^9$  IU. This total dose is approximately half the highest dose proposed for the human study. When considered on a IU/kg basis, a 6 kg monkey received a dose approximately 3 times greater than the highest proposed dose for a 60 kg human.



Timing of evaluations.

The animals were evaluated on the day of administration, and on days 3, 7, 24, 38, and 44 days after infection. The second administration of virus occurred on day 44. The monkeys were evaluated on day 48 and then on days 55, 62, and 129.

- 5 For evaluations, monkeys were anesthetized by intramuscular injection of ketamine (15 mg/kg). To obtain nasal epithelial cells after the first viral administration, the nasal mucosa was first impregnated with 5 drops of Afrin (0.05% oxymetazoline hydrochloride, Schering-Plough) and 1 ml of 2% Lidocaine for 5 minutes. A cytobrush was then used to gently rub the mucosa for about 3 sec. To obtain pharyngeal epithelial swabs, a cotton-tipped applicator was rubbed over the back of the pharynx 2-3 times. The resulting cells were dislodged from brushes or applicators into 2 ml of sterile PBS. After the second administration of Ad2-ORF6/PGK-CFTR, the monkeys were followed clinically for 3 weeks, and mucosal biopsies were obtained from the monkeys medial turbinate at days 4, 11 and 18.

15 Animal evaluation.

Animals were evaluated daily for evidence of abnormal behavior of physical signs. A record of food and fluid intake was used to assess appetite and general health. Stool consistency was also recorded to check for the possibility of diarrhea. At each of the evaluation time points, rectal temperature, respiratory rate, and heart rate were measured.

- 20 The nasal mucosa, conjunctivas and pharynx were visually inspected. The monkeys were also examined for lymphadenopathy.

Hematology and serum chemistry

Venous blood from the monkeys was collected by standard venipuncture technique.

- 25 Blood/serum analysis was performed in the clinical laboratory of the University of Iowa Hospitals and Clinics using a Hitachi 737 automated chemistry analyzer and a Technicom H6 automated hematology analyzer.

Serology

- 30 Sera from the monkeys were obtained and anti-adenoviral antibody titers were measured by ELISA. For the ELISA, 50 ng/well of killed adenovirus (Lee Biomolecular Research Laboratories, San Diego, Ca) was coated in 0.1M NaHCO<sub>3</sub> at 4° C overnight on 96 well plates. The test samples at appropriate dilutions were added, starting at a dilution of 1/50. The samples were incubated for 1 hour, the plates washed, and a goat anti-human IgG HRP conjugate (Jackson ImmunoResearch Laboratories, West Grove, PA) was added for 1 hour. The plates were washed and O-Phenylenediamine (OPD) (Sigma Chemical Co., St. Louis, MO) was added for 30 min. at room temperature. The assay was stopped with 4.5 M H<sub>2</sub>SO<sub>4</sub> and read at 490 nm on a Molecular Devices microplate reader. The titer was calculated as the product of the reciprocal of the initial dilution and the reciprocal of the

dilution in the last well with an OD>0.100. Nasal washings from the monkeys were obtained and anti-adenoviral antibody titers were measured by ELISA, starting at a dilution of 1/4.

#### Nasal Washings.

- 5 Nasal washings were obtained to test for the possibility of secretory antibodies that could act as neutralizing antibodies. Three ml of sterile PBS was slowly instilled into the nasal cavity of the monkeys, the fluid was collected by gravity. The washings were centrifuged at 1000 RPM for 5 minutes and the supernatant was used for anti-adenoviral, and neutralizing antibody measurement.

10

#### Cytology

- Cells were obtained from the monkey's nasal epithelium by gently rubbing the nasal mucosa for about 3 seconds with a cytobrush. The resulting cells were dislodged from the brushes into 2 ml of PBS. The cell suspension was spun at 5000 rpm for 5 min. and  
15 resuspended in 293 media at a concentration of  $10^6$  cells/ml. Forty  $\mu$ l of the cell suspension was placed on slides using a Cytospin. Cytospin slides were stained with Wright's stain and analyzed for cell differential using light microscopy.

#### Culture for Ad2-ORF6/PFK-CFTR

- 20 To assess for the presence of infectious viral particles, the supernatant from the nasal brushings and pharyngeal swabs of the monkeys were used. Twenty-five  $\mu$ l of the supernatant was added in duplicate to 293 cells. 293 cells were used at 50% confluence and were seeded in 96 well plates. 293 cells were incubated for 72 hours at 37°C, then fixed with a mixture of equal parts of methanol and acetone for 10 min and incubated with an FITC  
25 label anti-adenovirus monoclonal antibodies (Chemicon, Light Diagnostics, Temecuca, Ca) for 30 min. Positive nuclear immunofluorescence was interpreted as positive culture.

#### Immunocytochemistry for the detection of CFTR.

- Cells were obtained by brushing. Eighty  $\mu$ l of cell suspension were spun onto gelatin-coated slides. The slides were allowed to air dry, and then fixed with 4% paraformaldehyde. The cells were permeabilized with 0.2 Triton-X (Pierce, Rockford, Il) and then blocked for 60 minutes with 5% goat serum (Sigma, Mo). A pool of monoclonal antibodies (M13-1, M1-4, and M6-4) (Gregory et al., (1990) *Nature* 347:382-386); Denning et al., (1992) *J. Cell Biol.* 118:(3) 551-559; Denning et al., (1992) *Nature* 358:761-764) were added and incubated for  
35 12 hours. The primary antibody was washed off and an antimouse biotinylated antibody (Biomed, Foster City, Ca) was added. After washing, the secondary antibody, streptavidin FITC (Biomed, Foster City, Ca) was added and the slides were observed with a laser scanning confocal microscope.

### Biopsies

- To assess for histologic evidence of safety, nasal medial turbinate biopsies were obtained on day 4, 11 and 18 after the second viral administration as described before (Zabner et al (1993) Human Gene Therapy, in press). Nasal biopsies were fixed in 4% formaldehyde and H&E stained sections were reviewed.

## **RESULTS**

### Studies of efficacy.

- To directly assess the presence of CFTR, cells obtained by brushing were plated onto slides by cytopsin and stained with antibodies to CFTR. A positive reaction is clearly evident in cells exposed to Ad2-ORF6/PGK-CFTR. The cells were scored as positive by immunocytochemistry when evaluated by a reader blinded to the identity of the samples. Cells obtained prior to infection and from other untreated monkeys were used as negative controls. Figures 36A-36D, 37A-37D, and 38A-38D show examples from each monkey.

### Studies of safety

- None of the monkeys developed any clinical signs of viral infections or inflammation.
- There were no visible abnormalities at days 3, 4, 7 or on weekly inspection thereafter. Physical examination revealed no fever, lymphadenopathy, conjunctivitis, coryza, tachypnea, or tachycardia at any of the time points. There was no cough, sneezing or diarrhea. The monkeys had no fever. Appetites and weights were not affected by virus administration in either monkey. The data are summarized in Figures 39A-39C.
- The presence of live virus was tested in the supernatant of cell suspensions from swabs and brushes from each nostril and the pharynx. Each supernatant was used to infect the virus-sensitive 293 cell line. Live virus was never detected at any of the time points. The rapid loss of live virus suggests that there was no viral replication.
- The results of complete blood counts, sedimentation rate, and clinical chemistries are shown in Figure 40A-40C. There was no evidence of a systemic inflammatory response or other abnormalities of the clinical chemistries.
- Epithelial inflammation was assessed by cytological examination of Wright-stained cells (cytopsin) obtained from brushings of the nasal epithelium. The percentage of neutrophils and lymphocytes from the infected nostrils were compared to those of the control nostrils and values from four control monkeys. Wright stains of cells from nasal brushing were performed on each of the evaluation days. Neutrophils and lymphocytes accounted for less than 5% of total cells at all time points. The data are shown in Figure 41. The data indicate that administration of Ad2-ORF6/PGK-CFTR caused no change in the distribution or number of inflammatory cells at any of the time points following virus administration,

even during a second administration of the virus. The biopsy slides obtained after the second Ad2-ORF6/PGK-CFTR administration were reviewed by an independent pathologist, who found no evidence of inflammation or any other cytopathic effects. Figures 42 to 44 show an example from each monkey.

5        Figures 45A-45C shows that all three monkeys had developed antibody titers to adenovirus prior to the first infection with Ad2-ORF6/PGK-CFTR (Zabner et al. (1993) *Human Gene Therapy* (in press)). Antibody titers measured by ELISA rose within one week after the first and second administration and peaked at day 24. No anti-adenoviral antibodies were detected by ELISA or neutralizing assay in nasal washings of any of the monkeys.

10        These results combined with demonstrate the ability of a recombinant adenovirus encoding CFTR (Ad2-ORF6/PGK-CFTR) to express CFTR cDNA in the airway epithelium of monkeys. These monkeys have been followed clinically for 12 months after the first viral administration and no complications have been observed.

15        The results of the safety studies are encouraging. No evidence of viral replication was found; infectious viral particles were rapidly cleared. The other major consideration for safety of an adenovirus vector in the treatment of CF is the possibility of an inflammatory response. The data indicate that the virus generated an antibody response, but despite this, no evidence of a systemic or local inflammatory response was observed. The cells obtained by brushings and swabs were not altered by virus application. Since these Monkeys had been  
20        previously exposed three times to Ad2/CFTR-1, these data suggest that at least five sequential exposures of airway epithelium to adenovirus does not cause a detrimental inflammatory response.

25        These data indicate that Ad2-ORF6/PGK-CFTR can effectively transfer CFTR cDNA to airway epithelium and direct the expression of CFTR. They also indicate that transfer and expression is safe in primates.

#### Equivalents

30        Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

**TABLE I**

<b>Mutant</b>	<b>CF</b>	<b>Exon</b>	<b>CFTR Domain</b>	<b>A</b>	<b>B</b>
Wild Type				-	+
R334W	Y	7	TM6	-	+
K464M	N	9	NBD1	-	+
Δ1507	Y	10	NBD1	-	+
ΔF508	Y	10	NBD1	-	+
F508R	N	10	NBD1	-	+
S549I	Y	11	NBD1	-	+
G551D	Y	11	NBD1	-	+
N894,900Q	N	15	ECD4	+	-
K1250M	N	20	NBD2	-	+
Tth111	N	22	NB-Term	-	+

Table II.

10	20	30	40	50	60
CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT					
GTAGTAGTTA TTATATGGAA TAAACCTAA CTTCGGTTAT ACTATTACTC CCCCACCTCA					
____INVERTED TERMINAL REPETITION-ORIGIN OF REPLICATION____					60>
70	80	90	100	110	120
TTGTGACGTG GCGCGGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCGGAAGTGT					
AACACTGCAC CGCGCCCCGC ACCCTTGCCC CGCCCACTGC ATCATCACAC CGCCTTCACA					
____INVERTED TERMINAL REPETITION-ORIGIN OF R____>					
130	140	150	160	170	180
GATGTTGCAA GTGTGGCGGA ACACATGTAA GCGCCGGATG TGGTAAAAGT GACGTTTTTG					
CTACAACGTT CACACCGCCT TGTGTACATT CGCGGCCTAC ACCATTTTCA CTGCAAAAAC					
190	200	210	220	230	240
GTGTGCGCCG GTGTATACGG GAAGTGACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG					
CACAGCGCGC CACATATGCC CTTCACTGTT AAAAGCGCGC CAAATCCGC CTACAACATC					
____b____E1A ENHANCER AND VIRAL PACKAGING DOMAIN____					50>
250	260	270	280	290	300
TAAATTTGGG CGTAACCAAG TAATGTTTGG CCATTTTCGC GGGAAACTG AATAAGAGGA					
ATTAAACCC GCATTGGTTC ATTACAAACC GGTAAAGCG CCTTTTGAC TTATCTCCT					
____60_b_E1A ENHANCER AND VIRAL PACKAGING DOMAIN_0_b____					110>
310	320	330	340	350	360
AGTGAAATCT GAATAATTCT GTGTACTCA TAGCGGTAA TATTTGTCTA GGGCCGCGGG					
TCACTTTAGA CTTATTAAGA CACAATGAGT ATCGCGCATT ATAAACAGAT CCGCGCGCCC					
____120_b_E1A ENHANCER AND VIRAL PACKAGING DOMAIN_0_b____					170>
370	380	390	400	410	420
GACTTTGACC GTTTACGTGG AGACTCGCCC AGGTGTTTT CTCAGGTGTT TTCCGCGTTC					
CTGAACTGG CAAATGCACC TCTGAGCGGG TCCACAAAA GAGTCCACAA AAGGCGCAAG					
____E1A ENHANCER a_90>					
____c____10_E1A PROMOTER REGION_0_c____					40>
430	440	450	460	470	480
CGGCTCAAAG TTGGCGTTTT ATTATTATAG TCAGCTGACC CGCAGTGTAT TTATACCCGG					
GCCAGTTTC AACCGCAAAA TAATAATATC AGTCGACTGC CCGTCACATA AATATGGGCC					
____50_c____60_E1A PROMOTER REGION_c____					90_c____100>
490	500	510	520	530	540
TGAGTTCCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTTCTCC TCCGAGCCGC					
ACTCAAGGAG TTCTCCGGTG AGAAGTCAGG GTCGCTCATC TCAAAAGAGG AGGCTCGGCG					
____h____HYBRID E1A-CFTR-E1B MESSAGE					
____E1A PROMOTER_120>					
____o____E1A RNA 5' UNTRANSLATED_o____					40>
550	560	570	580	590	600
TCCGAGCTAG TAACGGCCGC CAGTGCTCTC CAGATATCAA AGTCGACGGT ACCCGAGAGA					
AGGCTCGATC ATTGCCGGCG GTCACAGCAC GTCTATAGTT TCAGCTGCCA TGGGCTCTCT					

\_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_>  
 \_\_\_\_\_e\_\_\_\_\_10\_\_\_\_\_SYNTHETIC LINKER SEQUENCES\_\_\_\_\_40\_\_\_\_\_e\_\_\_\_\_>  
 \_\_\_\_\_130>

610 620 630 640 650 660  
 CCATGCAGAG GTCGCCTCTG GAAAAGGCCA GCGTTGTCTC CAAACTTTTT TTCAGCTGGA  
 GGTACGTCTC CAGCGGAGAC CTTTCGGGT CGCAACAGAG GTTTGAAAAA AAGTCGACCT  
 M Q R S P L E K A S V V S K L F F S W>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_140i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_180i\_\_\_\_\_190>

670 680 690 700 710 720  
 CCAGACCAAT TTTGAGGAAA GGATACAGAC AGCGCCTGGA ATTGTCAGAC ATATACCAAA  
 GGTCTGGTTA AAACCTCCTTT CCTATGTCTG TCGCGGACCT TAACAGTCTG TATATGGTIT  
 T R P I L R K G Y R Q R L E L S D I Y Q>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_200i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_240i\_\_\_\_\_250>

730 740 750 760 770 780  
 TCCCTTCTGT TGATTCTGCT GACAATCTAT CTGAAAAATT GGAAAGAGAA TGGGATAGAG  
 AGGGAAGACA ACTAAGACGA CTGTTAGATA GACTTTTTAA CCTTTCTCTT ACCCTATCTC  
 I P S V D S A D N L S E K L E R E W D R>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_260i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_300i\_\_\_\_\_310>

790 800 810 820 830 840  
 AGCTGGCTTC AAAGAAAAAT CCTAAACTCA TTAATGCCCT TCGGCGATGT TTTTCTGGA  
 TCGACCGAAG TTTCTTTTAA GGATTGAGT AATTACGGGA AGCCGCTACA AAAAAGACCT  
 E L A S K K N P K L I N A L R R C F F W>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_320i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_360i\_\_\_\_\_370>

850 860 870 880 890 900  
 GATTTATGTT CTATGGAATC TTTTATATT TAGGGGAAGT CACCAAAGCA GTACAGCCTC  
 CTAATACAA GATACCTTAG AAAATATAA ATCCCTTCA GTGGTTTCGT CATGTCGGAG  
 R F M F Y G I F L Y L G E V T K A V Q P>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_380i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_420i\_\_\_\_\_430>

910 920 930 940 950 960  
 TCTTACTGGG AAGATCATA GCTTCCTATG ACCCGGATAA CAAGGAGGAA CGCTCTATCG  
 AGRATGACCC TTCTTAGTAT CGAAGGATAC TGGGCCTATT GTTCCTCCTT GCGAGATAGC  
 L L L G R I I A S Y D P D N K E E R S I>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_440i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_480i\_\_\_\_\_490>

970 980 990 1000 1010 1020  
 CGATTTATCT AGGCATAGGC TTATGCCTTC TCTTTATTGT GAGGACACTG CTCCTACACC

GCTAAATAGA TCCGTATCCG AATACGGAAG AGAAATAACA CTCCTGTGAC GAGGATGTGG  
 A I Y L G I G L C L L F I V R T L L L H>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_500i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_540i\_\_\_\_550>

1030 1040 1050 1060 1070 1080  
 CAGCCATTTT TGGCCTTCAT CACATTGGAA TGCAGATGAG AATAGCTATG TTTAGTTTGA  
 GTCGGTAAAA ACCGGAAGTA GTGTAACCTT ACGTCTACTC TTATCGATAC AAATCAAAC  
 P A I F G L H H I G M Q H R I A M F S L>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_560i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_600i\_\_\_\_610>

1090 1100 1110 1120 1130 1140  
 TTTATAAGAA GACTTTAAAG CTGTCAAGCC GTGTTCTAGA TAAATAAGT ATTGGACAAC  
 AAATATCTTT CTGAAATTTT CACAGTTTCG CACAAGATCT ATTTTATTCA TAACCTGTTG  
 I Y K K T L K L S S R V L D K I S I G Q>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_620i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_660i\_\_\_\_670>

1150 1160 1170 1180 1190 1200  
 TTGTTAGTCT CCTTCCAAC AACCTGAACA AATTGATGA AGGACTTGCA TTGGCACATT  
 AACATCAGA GGAAAGGTTG TTGGACTTGT TAAACTACT TCCTGAACGT AACCGTGTAA  
 L V S L L S N N L N K F D E G L A L I A H>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_680i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_720i\_\_\_\_730>

1210 1220 1230 1240 1250 1260  
 TCGTGTGGAT CGCTCCTTTG CAAGTGGCAC TCCTCATGGG GCTAATCTGG GAGTTGTTAC  
 AGCACACCTA GCGAGGAAAC GTTCAACCGT AGGAGTACCC CGATTAGACC CTCAACAATG  
 F V W I A P L Q V A L L M G L I W E L L>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_740i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_780i\_\_\_\_790>

1270 1280 1290 1300 1310 1320  
 AGGCGTCTGC CTTCTGTGGA CTTGGTTTCC TGATAGTCCT TGCCCTTTTT CAGGCTGGGC  
 TCCGACGACG GAAGACACCT GAACCAAAGG ACTATCAGGA ACGGGAAGAA GTCCGACCCG  
 Q A S A F C G L G F L I V L A L F Q A G>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_800i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_840i\_\_\_\_850>

1330 1340 1350 1360 1370 1380  
 TAGGGAGAAT GATGATGAAG TACAGAGATC AGAGAGCTGG GAAGATCACT GAAAGACTTG  
 ATCCCTCTTA CTACTACTTC ATGTCTCTAG TCTCTCGACC CTTCTAGTCA CTTTCTGAAC  
 L G R M M M K Y R D Q R A G K I S E R L>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_860i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_900i\_\_\_\_910>

1390 1400 1410 1420 1430 1440



TGATTACCTC AGAAATGATT GAAACATCC AATCTGTAA GGCATACTGC TGGGAACAAG  
 ACTAATGGAG TCTTTACTAA CTTTGTAGG TTAGACAATT CCGTATGACG ACCCTTCTTC  
 V I T S E M I E N I Q S V K A Y C W E E  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_920i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_960i\_\_\_\_970>

1450 1460 1470 1480 1490 1500

CAATGGAAA AATGATTGAA AACTTAAGAC AAACAGAAGT GAACTGACT CGGAAGGCAG  
 GTTACCTTTT TTACTAAGTT TTGAATTCG TTGTCTTGA CTTTGACTGA GCCTTCCGTC  
 A M E K M I E N L R Q T E L K L T R K A  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_980i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1020i\_\_\_\_1030>

1510 1520 1530 1540 1550 1560

CCTATGTGAG ATACTTCAAT AGCTCAGCCT TCTTCTCTC AGGGTTCTTT GTGGTGTTTT  
 GGATACACTC TATGAAGTTA TCGAGTCGGA AGAAGAAGAG TCCCAAGAAA CACCACAAA  
 A Y V R Y F N S S A F F F S G F F V V F  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_1040i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1080i\_\_\_\_1090>

1570 1580 1590 1600 1610 1620

TATCTGTGCT TCCCTATGCA CTAATCAAG GAATCATCCT CCGGAAATA TTCACCA  
 ATAGACACGA AGGGATACGT GATTAGTTTC CTTAGTAGGA GGCCTTTTAT AAGTGGTGGT  
 L S V L P Y A L I K G I I L R K I F T T  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_1100i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1140i\_\_\_\_1150>

1630 1640 1650 1660 1670 1680

TCTCATTCTG CATTGTTCTG CGCATGGCGG TCACTCGGCA ATTTCCCTGG GCTGTACAAA  
 AGAGTAAGAC GTAACAAGAC GCGTACCGCC AGTGAGCCGT TAAAGGGACC CGACATGTTT  
 I S F C I V L R M A V T R Q F P W A V Q  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_1160i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1200i\_\_\_\_1210>

1690 1700 1710 1720 1730 1740

GATGSTATGA CTCTCTTGGG GATATTAACA AATACAGGA TTCTTTACAA AAGCAAGAA  
 GTACCATACT GAGAGAACCT CGTTATTTGT TTATGTCTCT AAAGAATGTT TTCGTTCTTA  
 T W Y D S L G A I N K I Q D F L Q K Q E  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_1220i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1260i\_\_\_\_1270>

1750 1760 1770 1780 1790 1800

AATAGACATT GGAATATAAC TTAACGACTA CAGAAGTAGT GATGGAGGAT GTAACAGCCT  
 TATTCTGTAA CCTTATATTG AATTGCTGAT GTCTTCATCA CTACCTCTTA CATTGTGCGA  
 Y K T L E Y N L T T T E V V M E N V T A  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_1280i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_1320i\_\_\_\_1330>

1810 1820 1830 1840 1850 1860

-71-

TCTGGGAGGA GGGATTTGGG GAATTATTG AGAAAGCAA ACIAAACAAT AACAAATAGAA  
 AGACCTCTCT CCTAAACCC CTTAATAAC TCTTCGTTT TGTTTTGTTA TTGTTATCTT  
 F W E E G F G E L F E K A K Q N N N N R>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1340i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1380i\_\_\_1390>

1870 1880 1890 1900 1910 1920

AAACTTCTAA TGGTGATGAC AGCCTCTTCT TCAGTAATTT CTCACTTCTT GGTACTCCTG  
 TTTGAAGATT ACCACTACTG TCGGAGAAGA AGTCATTAAA GAGTGAAGAA CCATGAGGAC  
 K T S N G D D S L F F S N F S L L G T P>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1400i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1440i\_\_\_1450>

1930 1940 1950 1960 1970 1980

TCTTGAAAGA TATTAATTTT AAGATAGAAA GAGGACAGTT GTTGGCGGTT GCTGGATCCA  
 AGGACTTTCT ATAATTAAAG TTCTATCTTT CTCCTGTCAA CAACCGCCAA CGACCTAGGT  
 V L K D I N F K I E R G Q L L A V A G S>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1460i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1500i\_\_\_1510>

1990 2000 2010 2020 2030 2040

CTGGAGCAGG CAAGACTTCA CTTCTAATGA TGATTATGGG AGAACTGGAG CCTTCAGAGG  
 GACCTCGTCC GTTCTGAAGT GAAGATTACT ACTAATACCC TCTTGACCTC GGAAGTCTCC  
 T G A G K T S L L M M I M G E L E P S E>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1520i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1560i\_\_\_1570>

2050 2060 2070 2080 2090 2100

GTAAAATTAA GCACAGTGGG AGAATTTTCT TCTGTTCTCA GTTTTCCTGG ATTATGCCTG  
 CATTTTAATT CGTGTACCTT TCTTAAAGTA AGACAAGAGT CAAAAGGACC TAATACGGAC  
 G K I K H S G R I S F C S Q F S W I M P>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1580i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1620i\_\_\_1630>

2110 2120 2130 2140 2150 2160

GCACATTAA AGAAATATC ATCTTTGGTG TTTCCTATGA TGAATATAGA TACAGAAGCG  
 CGTGGAATTT TCTTTTATAG TAGAAACCAC AAAGGATACT ACTTATATCT ATGTCTTCGC  
 G T I K E N I I F G V S Y D E Y R Y R S>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1640i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1680i\_\_\_1690>

2170 2180 2190 2200 2210 2220

TCATCAAGC ATGCCAACTA GAAGAGGACA TCTCCAAGTT TGCAGAGAAA GACAATATAG  
 AGTAGTTTCG TACGGTTGAT CTTCTCCTGT AGAGSTTCAA ACGTCTCTTT CTGTTATATC  
 V I K A C Q L E E D I S K F A E K D N I>

\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_>  
 \_\_\_h\_\_\_HYBRID ELA-CFTR-ELB MESSAGE \_\_\_h\_\_\_>  
 \_\_\_1700i\_\_\_123 TO 4622 OF HUMAN CFTR CDNA \_\_\_1740i\_\_\_1750>

-72-

2230	2240	2250	2260	2270	2280
TTCTTGGAGA AGGTGGAATC AACTGAGTG GAGGTCAACG AGCAAGAATT TCTTTAGCAA AAGAACCCTCT TCCACCTTAG TGTGACTCAC CTCAGATTGC TCGTTCTTAA AGAAATCGTT V L G E G G I T L S G G O R A R I S L A> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____1760i____123 TO 4622 OF HUMAN CFTR CDNA____1800i____1810>					
2290	2300	2310	2320	2330	2340
GAGCAGTATA CAAAGATGCT GATTTGTATT TATTAGACTC TCCTTTTGGG TACCTAGATG CTCGTCATAT GTTCTACGA CTAACATAA ATAATCTGAG AGGAAAACCT ATGGATCTAC R A V Y K D A D L Y L L D S P F G Y L D> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____1820i____123 TO 4622 OF HUMAN CFTR CDNA____1860i____1870>					
2350	2360	2370	2380	2390	2400
TTTTAACAGA AAAAGAAATA TTTGAAAGCT GTGTCTGTAA ACTGATGGCT AACAAAACCTA AAAATTGTCT TTTTCTTTAT AAACCTTCGA CACAGACATT TGACTACCGA TTGTTTGTAT V L T E K E I F E S C V C K L M A N K T> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____1880i____123 TO 4622 OF HUMAN CFTR CDNA____1920i____1930>					
2410	2420	2430	2440	2450	2460
GGATTTTGGT CACTTCTAAA ATGGAACATT TAAAGAAAGC TGACAAATA TTAATTTTGC CCTAAAACCA GTGAAGATTT TACCTGTAA ATTCTTTTCG ACTGTTTAT AATTAAAACG R I L V T S K M E H L K K A D K I L I L> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____1940i____123 TO 4622 OF HUMAN CFTR CDNA____1980i____1990>					
2470	2480	2490	2500	2510	2520
ATGAAGGTAG CAGCTATTTT TATGGGACAT TTTCAGAACT CCAAATCTA CAGCCAGACT TACTTCCATC GTCGATAAAA ATACCCTGTA AAAGTCTTGA GGTTTTAGAT GTCGGTCTGA H E G S S Y F Y G T F S E L Q N L Q P D> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____2000i____123 TO 4622 OF HUMAN CFTR CDNA____2040i____2050>					
2530	2540	2550	2560	2570	2580
TTAGCTCAAA ACTCATGGGA TGTGATTTCT TCGACCAATT TAGTGCAGAA AGAAGAAATT AATCGAGTTT TGAGTACCCT AACTAAGAA AGCTGTTTAA ATCAGCTCTT TCTTCTTTAA F S S K L M G C D S F D Q F S A E R R N> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____2060i____123 TO 4622 OF HUMAN CFTR CDNA____2100i____2110>					
2590	2600	2610	2620	2630	2640
CAATCCTAAC TGAGACCTTA CACCGTTTCT CATTAGAAGG AGATGCTCCT GTCTCCTGGA GTTAGGATTG ACTCTGGAAT GTGGCAAGA GTAATCTTCC TCTACGAGGA CAGAGGACCT S I L T E T L H R F S L E G D A P V S W> ____CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON____> ____h____HYBRID ELA-CFTR-ELB MESSAGE____h____> ____2120i____123 TO 4622 OF HUMAN CFTR CDNA____2160i____2170>					

-73-

2650 2660 2670 2680 2690 2700  
CAGAAACAAA AAAACAATCT TTAAACAGA CTGGAGAGTT TGGGGAAAAA AGGAAGAATT  
GTCTTTGTTT TTTTGTAGA AAATTTGTCT GACCTCTCAA ACCCCTTTT TCCTTCTTAA  
T E T K K Q S F K Q T G E F G E K R K N>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2180i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2220i\_\_\_\_ 2230>  
2710 2720 2730 2740 2750 2760  
CTATTCTCAA TCCAATCAAC TCTATACGAA AATTTTCCAT TGTGCAAAAG ACTCCCTTAC  
GATAAGAGTT AGGTTAGTTG AGATATGCTT TTAAAAGGTA ACACGTTTTC TGAGGGAATG  
S I L N P I N S I R K F S I V Q K T P L>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2240i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2280i\_\_\_\_ 2290>  
2770 2780 2790 2800 2810 2820  
AAATGAATGG CATCGAAGAG GATTCTGATG AGCCTTTTGA GAGAAGGCTG TCCTTAGTAC  
TTTACTTACC GTAGCTTCTC CTAAGACTAC TCGGAATCT CTCTTCCGAC AGGAATCATG  
Q M N G I E E D S D E P L E R R L S L V>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2300i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2340i\_\_\_\_ 2350>  
2830 2840 2850 2860 2870 2880  
CAGATTCTGA GCAGGGAGAG GCGATACTGC CTCGCATCAG CGTGATCAGC ACTGGCCCCA  
GTCTAAGACT CGTCCCTCTC CGCTATGACG GAGCGTAGTC GCACTAGTCG TGACCGGGGT  
P D S E Q G E A I L P R I S V I S T G P>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2360i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2400i\_\_\_\_ 2410>  
2890 2900 2910 2920 2930 2940  
CGCTTCAGGC ACCAAGGAGG CAGTCTGTCC TGAACCTGAT GACACACTCA GTTAACCAAG  
CGGAAGTCCG TGCTTCTTCC GTCAGACAGG ACTTGGACTA CTGTGTGAGT CAATTGGTTC  
T L Q A R R R Q S V L N L M T H S V N Q>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2420i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2460i\_\_\_\_ 2470>  
2950 2960 2970 2980 2990 3000  
GTCAGACAT TCACCGAAG ACAACAGCAT CCACACGAAA AGTGTCACTG GCCCCTCAGG  
CAGTCTTGTA AGTGGCTTTC TGTGTCCTA GGTGTGCTTT TCACAGTGAC CGGGAGTCC  
G Q N I H R K T T A S T R K V S L A P Q>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>  
\_\_\_\_ 2480i\_\_\_\_ 123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_ 2520i\_\_\_\_ 2530>  
3010 3020 3030 3040 3050 3060  
CAACTTGAC TGAAGTGGAT ATATATTCAA GAGGTTATC TCAAGAACT GGCTTGGAAA  
GTTTGAAGT ACTTGACCTA TATATAAGTT CTTCCAATAG AGTTCTTTGA CCGAACCTTT  
A N L T E L D I Y S R R L S Q E T G L E>  
\_\_\_\_ CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
\_\_\_\_ h\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_>

\_\_\_\_\_2540i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2580i\_\_\_\_\_2590>  
 3070            3080            3090            3100            3110            3120  
 TAAGTGAAGA AATTAACGAA GAAGACTTAA AGGAGTGCCT TTTTGATGAT ATGGAGAGCA  
 ATTCACCTCT TTAATTGCTT CTTCTGAATT TCCTCACGGA AAAACTACTA TACCTCTCGT  
 I S E E I N E E D L K E C L F D D M E S>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2600i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2640i\_\_\_\_\_2650>  
 3130            3140            3150            3160            3170            3180  
 TACCAGCAGT GACTACATGG AACACATACC TTCGATATAT TACTGTCCAC AAGAGCTTAA  
 ATGGTCGTCA CTGATGTACC TTGTGTATGG AAGCTATATA ATGACAGGTG TTCTCGAATT  
 I P A V T T W N T Y L R Y I T V H K S L>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2660i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2700i\_\_\_\_\_2710>  
 3190            3200            3210            3220            3230            3240  
 TTTTGTGCT AATTGGTGC TTAGTAATTT TTCTGGCAGA GGTGGCTGCT TCTTTGGTTG  
 AAAAACACGA TTAACCACG AATCATTAA AGACCGTCT CCACCGACGA AGAAACCAAC  
 I F V L I W C L V I F L A E V A A S L V>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2720i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2760i\_\_\_\_\_2770>  
 3250            3260            3270            3280            3290            3300  
 TGCTGTGGCT CCTTGGAAAC ACTCCTCTC AAGACAAAGG GAATAGTACT CATAGTAGAA  
 ACACACCGA GGAACCTTTG TGAGGAGAAG TTCTGTTTCC CTTATCATGA GTATCATCTT  
 V L W L L G N T P L Q D K G N S T H S R>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2780i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2820i\_\_\_\_\_2830>  
 3310            3320            3330            3340            3350            3360  
 ATAACAGCTA TGCAGTGATT ATCACCAGCA CCAGTTCGTA TTATGTGTTT TACATTTACG  
 TATTGTCGAT ACGTCACTAA TAGTGGTCGT GGTCAGCAT AATACACAAA ATGTAAATGC  
 N N S Y A V I I T S T S S Y Y V F Y I Y>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2840i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2880i\_\_\_\_\_2890>  
 3370            3380            3390            3400            3410            3420  
 TGGGAGTAGC CGACACTTTG CTTGCTATGG GATTCTTCAG AGGTCTACCA CTGGTGCTA  
 AACTTCATCG GCTGTGAAC GAACGATACC CTAAGAAGTC TCCAGATGCT GACCACTAT  
 V G V A D T L L A M G F F R G L P L V M>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID ELA-CFTR-ELB MESSAGE\_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_2900i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_2940i\_\_\_\_\_2950>  
 3430            3440            3450            3460            3470            3480  
 CTCTAATCAC AGTGTGAAA ATTTTACACC ACAAATGTT ACATTCTGTT CTTCAAGCAC  
 GAGATTAGTG TCACAGCTTT TAAATGTGG TGTTTTACAA TGTAAGACAA GAAGTTTCGTG  
 T L I T V S K I L H H K M L H S V L Q A>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_>

h HYBRID ELA-CFTR-E1B MESSAGE h  
 2960i 123 TO 4622 OF HUMAN CFTR CDNA 3000i 3010>  
 3490 3500 3510 3520 3530 3540  
 CTATGTCAAC CCTCAACAGG TTGAAAGCAG GTGGGATTCT TAATAGATTTC TCCAAGATA  
 GATACAGTTG GGAGTTGTGC AACTTTCGTC CACCCTAAGA ATTATCTAAG AGGTTTCTAT  
 P M S T L N T L K A G G I L N R F S K D>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3020i 123 TO 4622 OF HUMAN CFTR CDNA 3060i 3070>  
 3550 3560 3570 3580 3590 3600  
 TAGCAATTTT GGATGACCTT CTGCCTCTTA CCATATTTGA CTTTCATCCAG TTGTTATTAA  
 ATCGTTAAAA CCTACTGGAA GACGGAGAAT GGTATAAACT GAAGTAGGTC AACAATAATT  
 I A I L D D L L P L T I F D F I Q L L L>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3080i 123 TO 4622 OF HUMAN CFTR CDNA 3120i 3130>  
 3610 3620 3630 3640 3650 3660  
 TTGTGATTGG AGCTATAGCA GTTGTGCGAG TTTTACAACC CTACATCTTT GTTGCAACAG  
 AACACTAACC TCGATATCGT CAACAGCGTC AAAATGTTGG GATGTAGAAA CAACGTTGTC  
 I V I G A I A V V A V L Q P Y I F V A T>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3140i 123 TO 4622 OF HUMAN CFTR CDNA 3180i 3190>  
 3670 3680 3690 3700 3710 3720  
 TGCCAGTGAT AGTGGCTTTT ATTATGTTGA GAGCATATTT CCTCCAAACC TCACAGCAAC  
 ACGGTCACTA TCACCGAAAA TAATACAACCT CTCGTATAAA GGAGGTTTGG AGTGTGCTTG  
 V P V I V A F I M L R A Y F L Q T S Q Q>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3200i 123 TO 4622 OF HUMAN CFTR CDNA 3240i 3250>  
 3730 3740 3750 3760 3770 3780  
 TCAAACAACT GGAATCTGAA GGCAGGAGTC CAATTTTCAC TCATCTTGTT ACAAGCTTAA  
 AGTTTGTGTA CCTTAGACTT CCGTCCTCAG GTTAAAAGTG AGTAGAACAA TGTTCGAATT  
 L K Q L E S E G R S P I F T H L V T S L>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3260i 123 TO 4622 OF HUMAN CFTR CDNA 3300i 3310>  
 3790 3800 3810 3820 3830 3840  
 AAGGACTATG GACACTTCGT GCCTTCGGAC GGCAGCCTTA CTTTGAAACT CTGTCCACA  
 TTCCTGATAC CTGTGAAGCA CGGAAGCCTG CCGTCGGAAT GAAACTTTGA GACAAGGTGT  
 K G L W T L R A F G R Q P Y F E T L F H>  
 CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON>  
 h HYBRID ELA-CFTR-E1B MESSAGE h  
 3320i 123 TO 4622 OF HUMAN CFTR CDNA 3360i 3370>  
 3850 3860 3870 3880 3890 3900  
 AAGCTCTGAA TTTACATACT GCCAACTGGT TCTTGACCT GTCAACACTG CGCTGTTTCC  
 TTCGAGACTT AATGTATGA CGGTTGACCA AGAATATGGA CAGTTGTGAC GCGACCAAGG  
 K A L N L H T A N W F L Y L S T L R W F>

-76-

\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3380i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3420i\_\_\_\_3430>

3910      3920      3930      3940      3950      3960  
 AAATGAGAAT AGAAATGATT TTTGTCATCT TCTTCATTGC TGTACCTTC ATTTCCATTT  
 TTTACTCTTA TCTTTACTAA AAACAGTAGA AGAAGTAACG ACAATGGAAG TAAAGGTAAA  
 Q M R I E M I F V I F F I A V T F I S I>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3440i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3480i\_\_\_\_3490>

3970      3980      3990      4000      4010      4020  
 TAACAACAGG AGAAGGAGAA GGAAGAGTTG GTATTATCCT GACTTTAGCC ATGAATATCA  
 ATTGTTGTCC TCTTCCTCTT CCTTCTCAAC CATAATAGGA CTGAAATCGG TACTTATAGT  
 L T T G E G E G R V G I I L T L A M N I>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3500i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3540i\_\_\_\_3550>

4030      4040      4050      4060      4070      4080  
 TGAGTACATT GCAAGTGGCT GTAACTCCA GCATAGATGT GGATAGCTTG ATGCGATCTG  
 ACTCATGTAA CGTCACCCGA CATTGAGGT CGTATCTACA CCTATCGAAC TACGCTAGAC  
 M S T L Q W A V N S S I D V D S L M R S>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3560i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3600i\_\_\_\_3610>

4090      4100      4110      4120      4130      4140  
 TGAGCCGAGT CTTTAAGTTC ATTGACATGC CAACAGAAGG TAAACCTACC AAGTCAACCA  
 ACTCGGCTCA GAAATTAAG TAACTGTACG GTTGTCTTCC ATTTGGATGG TTCAGTTGGT  
 V S R V F K F I D M P T E G K P T K S T>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3620i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3660i\_\_\_\_3670>

4150      4160      4170      4180      4190      4200  
 AACCATACAA GAATGGCCAA CTCTCGAAG TTATGATTAT TGAGAATTCA CACGTGAAGA  
 TTGGTATGTT CTTACCGGTT GAGAGCTTTC AATACTAATA ACTCTTAAGT GTGCACTTCT  
 K P Y K N G Q L S K V M I I E N S H V K>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3680i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3720i\_\_\_\_3730>

4210      4220      4230      4240      4250      4260  
 AAGATGACAT CTGGCCCTCA GGGGGCCAAA TGACTGTCAA AGATCTCACA GCAAAATACA  
 TTCTACTGTA GACCGGGAGT CCCCCGGTTT ACTGACAGTT TCTAGAGTGT CGTTTATGT  
 K D D I W P S G G Q M T V K D L T A K Y>  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3740i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3780i\_\_\_\_3790>

4270      4280      4290      4300      4310      4320  
 CAGAAGGTGG AAATGCCATA TTAGAGAAC TTTCTTCTC AATAAGTCCT GGGCAGAGGC  
 GTCTCCACC TTTACGGTAT AATCTCTCT AAAGGAAGAG TTATTCAGGA CCGGTCTCCC

-77-

T E G G N A I L E N I S F S I S P G Q R >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3800i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3840i\_\_\_\_3850>  
  
 4330      4340      4350      4360      4370      4380  
 TGGGCTCTT GGAAGAACT GGATCAGGA AGAGTACTTT GTTATCAGCT TTTTGTGAGAC  
 ACCCGGAGAA CCTTCTTGA CCTAGTCCCT TCTCATGAAA CAATAGTCGA AAAAAGCTCTG  
 V G L L G R T G S G K S T L L S A F L R >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3860i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3900i\_\_\_\_3910>  
  
 4390      4400      4410      4420      4430      4440  
 TACTGAACAC TGAAGGAGAA ATCCAGATCG ATGGTGTGTC TTGGGATTCA ATAACITTCG  
 ATGACTTGTG ACTTCCTCTT TAGGTCTAGC TACCACACAG AACCTAAGT TATTGAAACG  
 L L N T E G E I Q I D G V S W D S I T L >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3920i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_3960i\_\_\_\_3970>  
  
 4450      4460      4470      4480      4490      4500  
 AACAGTGGAG GAAAGCCTTT GGAGTGATAC CACAGAAAGT ATTTATTTTT TCTGGAACAT  
 TTGTCACCTC CTTTCGGAAA CCTCACTATG GTGCTTTCA TAAATAAAAA AGACCTTGTA  
 Q Q W R K A F G V I P Q K V F I F S G T >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_3980i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_4020i\_\_\_\_4030>  
  
 4510      4520      4530      4540      4550      4560  
 TTAGAAAAA CTGGATCCC TATGAACAGT GGAGTGATCA AGAAATATGG AAAGTTGCAG  
 AATCTTTTT GAACCTAGGG ATACTGTCA CTCCTAGT TCTTTATACC TTTCAACGTC  
 F R K N L D P Y E Q W S D Q E I W K V A >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_4040i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_4080i\_\_\_\_4090>  
  
 4570      4580      4590      4600      4610      4620  
 ATGAGGTTGG GCTCAGATCT GTGATAGAAC AGTTTCCTGG GAAGCTTGAC TTTGTCCTTG  
 TACTCCAACC CGAGTCTAGA CACTATCTTG TCAAAGGACC CTTGGAAGTG AAACAGGAC  
 D E V G L R S V I E Q F P G K L D F V L >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_4100i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_4140i\_\_\_\_4150>  
  
 4630      4640      4650      4660      4670      4680  
 TGGATGGGGG CTGTGCTCTA AGCCATGGCC ACGAGCAGTT GATGTGCTTG GCTAGATCTG  
 ACCTACCCCG GACACAGSAT TCGGTACCGG TGTTCTGCAA CTACACGAAC CGATCTAGAC  
 V D G G C V L S H G H K Q L M C L A R S >  
 \_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_>  
 \_\_\_\_h\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_h\_\_\_\_>  
 \_\_\_\_4160i\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_4200i\_\_\_\_4210>  
  
 4690      4700      4710      4720      4730      4740  
 TTCTCAGTAA GCGGAGATC TTGCTGCTG ATGATTCAG TGTCTATTC GATCCAGTAA



AAGAGTCATT CCGCTTCTAG AACGACGAAC TACTTGGGTC ACGAGTAAAC CTAGGTCATT  
 V L S K A K I L L D E P S A H L D P V>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4220i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4260i\_\_\_\_\_4270>

4750 4760 4770 4780 4790 4800

CATACCAAT AATTAGAAGA ACTCTAAAC AAGCATTTCG TGATTGCACA GTAATTCTCT  
 GTATGGTTTA TTAATCTTCT TGAGATTTCG TCGTAAACG ACTAACGTGT CATTAAGAGA  
 T Y Q I I R R T L K Q A F A D C T V I L>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4280i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4320i\_\_\_\_\_4330>

4810 4820 4830 4840 4850 4860

GTGAACACAG GATAGAAGCA ATGCTGGAAT GCCAACAATT TTTGGTCATA GAAGAGAACA  
 CACTTGTGTC CTATCTTCGT TAGGACCTTA CGGTGTGTA AAACCAAGTAT CTTCTCTTGT  
 C E H R I E A M L E C Q Q F L V I E E N>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4340i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4380i\_\_\_\_\_4390>

4870 4880 4890 4900 4910 4920

AAGTGGCGCA GTACGATTCC ATCCAGAAAC TGCTGAACGA GAGGAGCCTC TTCCGGCAAG  
 TTCACGCCGT CATGCTAAGG TAGGCTTTTG ACGACTTGCT CTCCTCGGAG AAGGCCGTTT  
 K V R Q Y D S I Q K L L N E R S L F R Q>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4400i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4440i\_\_\_\_\_4450>

4930 4940 4950 4960 4970 4980

CCATCAGCCC CTCCGACAGG GTGAAGCTCT TTCCCCACCG GAACTCAAGC AAGTGCAAGT  
 GGTAGTCGGG GAGGCTGTCC CACTTCGAGA AAGGGGTGGC CTTGAGTTCTG TTCACGTTCA  
 A I S P S D R V K L F P H R N S S K C K>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4460i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4500i\_\_\_\_\_4510>

4990 5000 5010 5020 5030 5040

CTAAGCCCCA GATTGCTGCT CTGAAGAGG AGACAGAAGA AGAGGTGCAA GATACAAGGC  
 GATTGGGGGT CTAACGACGA GACTTCTTCC TGTGCTTCT TCTCCACGTT CTATGTTCCG  
 S K P Q I A A L K E E T E E V Q D T R>  
 \_\_\_\_\_CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CODON\_\_\_\_\_  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_4520i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4560i\_\_\_\_\_4570>

5050 5060 5070 5080 5090 5100

TTTAGACAGC AGCATAAATG TTGACATGGG ACATTGCTC ATGGAATTGG AGGTAGCGGA  
 AATCTCTCTG TCGTATTTAC AACTGTACCC TGTAACGAG TACCTTAACC TCCATCGCCT  
 L >

\_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_\_  
 \_\_\_\_\_>

\_\_\_\_\_4580i\_\_\_\_\_123 TO 4622 OF HUMAN CFTR CDNA\_\_\_\_\_4620i\_\_\_\_\_>

-79-

5110 5120 5130 5140 5150 5160  
 TTGAGTACT GAAATGTGTG GCGGTGGCTT AAGGGTGGGA AAGAATATAT AAGGTGGGGG  
 AACTCCATGA CTTTACACAC CGCACCAGAA TTCCACCCCT TTCTTATATA TTCCACCCCC  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_10\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_50\_g\_\_\_\_\_60\_\_\_\_>  
 \_\_\_\_\_k\_\_\_\_\_10\_\_\_\_\_k\_\_\_\_\_ ELB 3' INTRON \_\_\_\_\_k\_\_\_\_\_40\_\_\_\_\_k\_\_\_\_\_50\_\_\_\_>

5170 5180 5190 5200 5210 5220  
 TCTCATGTAG TTTTGTATCT GTTTGTGAGC AGCCGCCGCC ATGAGCGCCA ACTCGTTTGA  
 AGAGTACATC AAAACATAGA CAAAACGTGC TCGCGGGCGG TACTCGCGGT TGAGCAAACCT  
 \_\_\_\_\_M\_\_\_\_\_S\_\_\_\_\_A\_\_\_\_\_N\_\_\_\_\_S\_\_\_\_\_F\_\_\_\_\_D\_\_\_\_\_>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_ IX MRNA \_\_\_\_\_l\_\_\_\_\_>  
 \_\_\_\_\_70\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_110\_g\_\_\_\_\_120\_\_\_\_>  
 \_\_\_\_\_60\_\_\_\_\_ ELB 3' INTRON \_\_\_\_\_80\_\_\_\_\_>

5230 5240 5250 5260 5270 5280  
 TGGAAGCATT GTGAGCTCAT ATTTGACAAC GCGCATGCCC CCATGGGCGG GGGTGGCTCA  
 ACCTTCGTAA CACTCGAGTA TAAACTGTG CCGGTACGGG GGTACCCGGC CCCACGCAGT  
 G S I V S S Y L T T R M P P W A G V R Q>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_ IX MRNA \_\_\_\_\_l\_\_\_\_\_>  
 \_\_\_\_\_130\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_170\_g\_\_\_\_\_180\_\_\_\_>

5290 5300 5310 5320 5330 5340  
 GAATGTGATG GGCTCCAGCA TTGATGGTCG CCCGTCCTG CCCGCAAACCT CTACTACCTT  
 CTTACACTAC CCGAGGTCGT AACTACCAGC GGGGCAGGAC GGGCGTTTGA GATGATGGAA  
 N V M G S S I D G R P V L P A N S T T L>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_ IX MRNA \_\_\_\_\_l\_\_\_\_\_>  
 \_\_\_\_\_190\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_230\_g\_\_\_\_\_240\_\_\_\_>

5350 5360 5370 5380 5390 5400  
 GACCTACGAG ACCGTGTCTG GAACGCCGTT GGAGACTGCA GCCTCCGCCG CCGCTTCAGC  
 CTGGATGCTC TGGCAGAGAC CTTGCGGCAA CCTCTGACGT CGGAGGCGGC GCGAAGTCG  
 T Y E T V S G T P L E T A A S A A A S A>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_ IX MRNA \_\_\_\_\_l\_\_\_\_\_>  
 \_\_\_\_\_250\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_290\_g\_\_\_\_\_300\_\_\_\_>

5410 5420 5430 5440 5450 5460  
 CGCTGCAGCC ACCGCCCGCG GGATTGTGAC TGACTTTGCT TTCCTGAGCC CGCTTGCAAG  
 GCGACGTCGG TGGCGGGCGC CCTACACTG ACTGAAACGA AAGGACTCGG GCGAACGTTT  
 A A A T A R G I V T D F A F L S P L A S>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_ HYBRID ELA-CFTR-ELB MESSAGE \_\_\_\_\_h\_\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_l\_\_\_\_\_1\_\_\_\_\_ IX MRNA \_\_\_\_\_l\_\_\_\_\_>  
 \_\_\_\_\_310\_g\_\_\_\_\_ ELB 3' UNTRANSLATED SEQUENCES \_\_\_\_\_350\_g\_\_\_\_\_360\_\_\_\_>

5470 5480 5490 5500 5510 5520  
 CAGTGCAGCT TCCCGTTTAT CCGCCCGCGA TGACAAGTTG ACGGCTCTTT TGGCACAATT

GTCACGTGGA AGGGCAAGTA GCGGGGCGCT ACTGTTCAAC TGCCGAGAAA ACCGTGTTAA  
 S A A S R S S A R D D K L T A L L A Q L>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_l\_\_\_\_\_IX MRNA\_\_\_\_\_l\_\_\_\_\_l\_\_\_\_>  
 \_\_\_\_\_370\_g\_\_\_\_\_E1B 3' UNTRANSLATED SEQUENCES\_\_\_\_\_410\_g\_\_\_\_\_420\_\_\_\_>

5530 5540 5550 5560 5570 5580

GGATTCTTTG ACCCGGGAAC TTAATGTCGT TTCTCAGCAG CTGTTGGATC TGCGCCAGCA  
 CCTAAGAAAC TGGGCCCTTG AATTACAGCA AAGAGTCGTC GACAACCTAG ACGCGGTCGT  
 D S L T R E L N V V S Q Q L L D L R Q Q>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); CODON\_START=1\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_l\_\_\_\_\_IX MRNA\_\_\_\_\_l\_\_\_\_\_l\_\_\_\_>  
 \_\_\_\_\_430\_g\_\_\_\_\_E1B 3' UNTRANSLATED SEQUENCES\_\_\_\_\_470\_g\_\_\_\_\_480\_\_\_\_>

5590 5600 5610 5620 5630

GGTTTCTGCC CTGAAGGCTT CCTCCCTCC CAATGCGGTT TAAAACATAA ATAAA  
 CCAAGACGG GACTTCCGAA GGAGGGGAGG GTTACGCCAA ATTTGTATT TATTT  
 V S A L K A S S P P N A V \*>  
 \_\_\_\_\_IX PROTEIN (HEXON-ASSOCIATED PROTEIN); C\_\_\_\_>  
 \_\_\_\_\_h\_\_\_\_\_HYBRID E1A-CFTR-E1B MESSAGE\_\_\_\_\_h\_\_\_\_>  
 \_\_\_\_\_l\_\_\_\_\_l\_\_\_\_\_IX MRNA\_\_\_\_\_l\_\_\_\_\_l\_\_\_\_>  
 \_\_\_\_\_490\_g\_\_\_\_\_E1B 3' UNTRANSLATED SEQUENCES\_\_\_\_\_530\_g\_\_\_\_\_>

Table III

## Nucleotide Sequence Analysis of Ad2-ORF6/PGK-CFTR

LOCUS	AD2-ORF6/P 36335 BP DS-DNA		
DEFINITION	-		
ACCESSION	-		
KEYWORDS	-		
SOURCE	-		
FEATURES	From	To/Span	Description
frag	12915	36335	10676 to 34096 of Ad2-E4/ORF6
frag	35069	35973	33178 to 34082 of Ad2 seq
pre-msg	> 35973	< 35069	(C) E4 mRNA [Nucleic Acids Res. 9, 1675-1689 (1981)], [J. Mol. Biol. 149, 189-221 (1981)], [Nucleic Acids Res. 12, 3503-3519 (1984)], [Unpublished (1984)] [Split]
IVS	35794	35084	(C) E4 mRNA intron D7 [J. Virol. 50, 106-117 (1984)], [Nucleic Acids Res. 12, 3503-3519 (1984)], [Unpublished (1984)]
IVS	35794	35175	(C) E4 mRNA intron D6 [Nucleic Acids Res. 12, 3503-3519 (1984)]
IVS	35794	35268	(C) E4 mRNA intron D5 [J. Virol. 50, 106-117 (1984)]
IVS	35794	35295	(C) E4 mRNA intron D4 [J. Virol. 50, 106-117 (1984)]
IVS	35794	35343	(C) E4 mRNA intron D3 [J. Virol. 50, 106-117 (1984)]
IVS	35794	35501	(C) E4 mRNA intron D2 [J. Virol. 50, 106-117 (1984)]
IVS	35794	35570	(C) E4 mRNA intron D1 [J. Virol. 50, 106-117 (1984)]
IVS	35794	35766	(C) E4 mRNA intron D [J. Virol. 50, 106-117 (1984)]
frag	35978	36335	35580 to 35937 of Ad2 seq
pre-msg	36007	< 35978	(C) E4 mRNA [Nucleic Acids Res. 9, 1675-1689 (1981)], [J. Mol. Biol. 149, 189-221 (1981)], [Nucleic Acids Res. 12, 3503-3519 (1984)], [Unpublished (1984)] [Split]
rpt	36234	36335	inverted terminal repetition; 99.54% [Biochem. Biophys. Res. Commun. 87, 671-678 (1979)], [J. Mol. Biol. 128, 577-594 (1979)]
frag	~ 12915	35054	1 to 32815 of Ad2 seq [Split]
pept	< 28478	28790	3 33K protein (virion morphogenesis)
pept	28478	28790	1 33K protein (virion morphogenesis); codon_start=1
mRNA	29331	< 12915	(C) E2b mRNA [J. Biol. Chem. 257, 13475-13491 (1982)] [Split]
pre-msg	< 12915	16352	major late mRNA L1 (alt.) [J. Mol. Biol. 149, 189-221 (1981)], [J. Virol. 48, 127-134 (1983)] [Split]
pre-msg	< 12915	20208	major late mRNA L2 (alt.) [J. Mol. Biol. 149, 189-221 (1981)], [J. Virol. 38, 469-482 (1981)], [J. Virol. 48, 127-134 (1983)] [Split]
pre-msg	< 12915	24682	major late mRNA L3 (alt.) [Nucleic Acids Res. 9, 1-17 (1981)], [J. Mol. Biol. 149, 189-221 (1981)], [J. Virol. 48, 127-134 (1983)] [Split]
pre-msg	< 12915	30462	major late mRNA L4 (alt.) [J. Mol. Biol. 149, 189-221 (1981)], [J. Virol. 48, 127-134 (1983)] [Split]
pre-msg	< 12915	35037	major late mRNA L5 (alt.) [J. Mol. Biol. 149, 189-221 (1981)], [J. Virol. 48, 127-134 (1983)] [Split]

## Nucleotide Sequence Analysis (cont.)

mRNA	< 12915	13278	major late mRNA intron (precedes 52,55K mRNA; 1st L1 mRNA) [Cell 16, 841-850 (1979)], [Cell 16, 851-861 (1979)], [J. Mol. Biol. 134, 143-158 (1979)], [J. Mol. Biol. 135, 413-433 (1979)], [Nature 292, 420-426 (1981)] [Split]
IVS	< 12915	16388	major late mRNA intron (precedes penton mRNA; 1st L2 mRNA) [J. Virol. 48, 127-134 (1983)] [Split]
IVS	< 12915	18754	major late mRNA intron (precedes pV mRNA; 2nd L2 mRNA) [J. Biol. Chem. 259, 13980-13985 (1984)] [Split]
IVS	< 12915	20238	major late mRNA intron (precedes pVI mRNA; 1st L3 mRNA) [J. Virol. 38, 469-482 (1981)] [Split]
IVS	< 12915	21040	major late mRNA intron (precedes hexon mRNA; 2nd L3 mRNA) [Proc. Natl. Acad. Sci. U.S.A. 75, 5822-5826 (1978)], [Cell 16, 841-850 (1979)] [Split]
IVS	< 12915	23888	major late mRNA intron (precedes 23K mRNA; 3rd L3 mRNA) [Nucleic Acids Res. 9, 1-17 (1981)] [Split]
IVS	< 12915	26333	major late mRNA intron (precedes 100K mRNA; 1st L4 mRNA) [Virology 128, 140-153 (1983)] [Split]
RNA	< 12915	13005	VA I RNA (alt.) [J. Biol. Chem. 252, 9043-9046 (1977)] [Split]
RNA	< 12915	13005	VA I RNA (alt.) [J. Biol. Chem. 246, 6991-7009 (1971)], [J. Biol. Chem. 252, 9047-9054 (1977)], [Proc. Natl. Acad. Sci. U.S.A. 77, 2424-2428 (1980)] [Split]
????	< 12915	13262	VA II RNA [Proc. Natl. Acad. Sci. U.S.A. 77, 3778-3782 (1980)], [Proc. Natl. Acad. Sci. U.S.A. 77, 2424-2428 (1980)] [Split]
pept	13279	14526	1 52,55K protein; codon_start=1
pept	14547	16304	1 IIIa protein (peripentonal hexon-associated protein; splice sites not sequenced); codon_start=1
signal	16331	16336	major late mRNA L1 poly-A signal (putative) 39.21%
pept	16390	18105	1 penton protein (virion component III); codon_start=1
pept	18112	18708	1 Pro-VII protein (precursor to major core protein); codon_start=1
pept	18778	19887	1 pV protein (minor core protein); codon_start=1
signal	20188	20193	major late mRNA L2 polyadenylation signal (putative) 49.94%
pept	20240	20992	1 pVI protein (hexon-associated precursor); codon_start=1
pept	21077	23983	1 hexon protein (virion component II); codon_start=1
????	< 12915	24631	23K protein (endopeptidase); codon_start=1 [Split]
signal	24657	24662	major late mRNA L3 polyadenylation signal (putative); 62.38%
pre-msg	28193	24659 (C)	E2a late mRNA (alt.) [J. Mol. Biol. 149, 189-221 (1981)]
pre-msg	28195	24659 (C)	E2a late mRNA (alt.) [Nucleic Acids Res. 12, 3503-3519 (1984)], [Unpublished (1984)]
pre-msg	29330	24659 (C)	E2a early mRNA (alt.) [J. Mol. Biol. 149,

-83-

## Nucleotide Sequence Analysis (cont.)

			189-221 (1981)]
pre-msg	29331	24659 (C)	E2a early mRNA (alt.) [J. Mol. Biol. 149, 189-221 (1981)]
signal	24683	24678 (C)	E2a mRNA polyadenylation signal on $\overline{\text{comp}}$ strand (putative); 62.43%
pept	26318	24729 (C)	DBP protein (DNA binding or 72K protein); codon_start=1
IVS	26953	26328 (C)	E2a mRNA intron B [Nucleic Acids Res. 9, 4439-4457 (1981)]
pept	26347	28764	1 100K protein (hexon assembly); codon_start=1
IVS	29263	27031 (C)	E2a early mRNA intron A [Cell 18, 569-580 (1979)]
IVS	28124	27211 (C)	E2a late mRNA intron A [Virology 128, 140-153 (1983)]
IVS	28791	28992	33K-pept intron [J. Virol. 45, 251-263 (1983)]
pept	28993	> 29366	1 33K protein (virion morphogenesis)
pept	29454	30137	1 pVIII protein (hexon-associated precursor); codon_start=1
mRNA	29848	33103	E3-2 mRNA; 85.88% [Gene 22, 157-165 (1983)]
IVS	30220	30614	major late mRNA intron ('x' leader) [Gene 22, 157-165 (1983)], [J. Biol. Chem. 259, 13980-13985 (1984)]
signal	30444	30449	major late mRNA L4 polyadenylation signal; (putative) 78.48%
signal	< 12915	32676	major late mRNA intron ('y' leader) [J. Mol. Biol. 135, 413-433 (1979)], [J. Virol. 38, 469-482 (1981)], [EMBO J. 1, 249-254 (1982)], [Gene 22, 157-165 (1983)] [Split]
pept	31051	31530	1 E3 19K protein (glycosylated membrane protein); codon_start=1
pept	31707	32012	1 E3 11.6K protein; codon_start=1
signal	32008	32013	E3-1 mRNA polyadenylation signal (putative); 82.69%
IVS	32822	33268	major late mRNA intron ('z' leader) [Proc. Natl. Acad. Sci. U.S.A. 75, 5822-5826 (1978)], [Cell 16, 841-850 (1979)], [EMBO J. 1, 249-254 (1982)], [Gene 22, 157-165 (1983)]
signal	33081	33086	E3-2 mRNA polyadenylation signal; 85.82% (putative)
????	< 12915	35017	fiber protein (virion component IV); codon_start=1 [Split]
signal	35013	35018	major late mRNA L5 polyadenylation signal; (putative) 91.19%
pre-msg	35054	> 35041 (C)	E4 mRNA [Nucleic Acids Res. 9, 1675-1689 (1981)], [J. Mol. Biol. 149, 189-221 (1981)], [Nucleic Acids Res. 12, 3503-3519 (1984)], [Unpublished (1984)] [Split]
frag	1	12914	1 to 12914 of pAd2/PGK-CPTR
DNA	1	> 356	1 to 357 Ad2
rpt	1	> 103	inverted terminal repetition; 0.28% [Biochem. Biophys. Res. Commun. 87, 671-678 (1979)], [J. Mol. Biol. 128, 577-594 (1979)]
	< 10	103	inverted terminal repetition; 0.28% [Biochem. Biophys. Res. Commun. 87, 671-678 (1979)], [J. Mol. Biol. 128, 577-594 (1979)] [Split]
frag	357	379	linker segment
frag	915	> 923	polylinker cloning sites [Split]

-84-

## Nucleotide Sequence Analysis (cont.)

DNA	< 924 > 954	polylinker cloning sites [Split]
signal	< 5567 > 12914	3328 to 10685 of Ad2 [Split]
frag	< 380 > 914	pgk promoter
frag	< 955 > 958	polylinker cloning sites [Split]
signal	< 5501 > 5522	polylinker cloning sites [Split]
frag	5523 > 5555	syn. BGH poly A
frag	5555 > 5560	linker [Split]
frag	< 5564 > 5567	linker [Split]
revision	959 5500	920 to 5461 of pCMV-CFTR-936C
modified	2868 2868	mistake in published sequence of Riordan et al. C not A is correct = N to H a.a. change
site	< 959 > 975	936 T to C mutation to inactivate cryptic bacterial promoter. Silent amino acid change
site	976 990	polylinker segment from pCMV-CFTR-936C (Rc/CMV-Invitrogen SpeI-BstXI) [Split]
site	991 1001	linker segment from pCMV-CFTR-936C. Originally Sali/BstXI adaptor oligo 1499DS
mRNA	1001 > 5500	linker segment from pCMV-CFTR-936C. Originally from pMT-CFTR construction oligo 1247 RG -Sal I to Aval sites.
pept	1011 > 5453	123 to 4622 of HUMCFTR
BASE COUNT	8597 A 10000 C 9786 G 7952 T 0 OTHER	cystic fibrosis transmembrane conductance regulator; codon_start=1
ORIGIN	?	

Ad2-ORF6/P Length: 36335 Sep 16, 1993 - 08:13 PM Check: 1664 ..

```

1 CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
61 TTGTGACGTG GCGCGGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCGGAAGTGT
121 GATGTTGCAA GTGTGGGGGA ACACATGTAA GCGCCGGATG TGGTAAAAGT GACGTTTTTG
181 GTGTGGGCGG GTGTATACGG GAAGTGACAA TTTTGGGGCG GTTTTAGGCG GATGTTGTAG
241 TAAATTTGGG CGTAACCAAG TAATGTTTGG CCATTTTTCG GGGAAAACGT AATAAGAGGA
301 AGTGAAATCT GAATAATTCT GTGTTACTCA TAGCGGTAA TATTGTCTTA GGGCGGCTCG
361 AGGTGACGGG TCTATGATA AGCTTGATAT CGAATTCGGG GGTGCGGGTT GCGCCTTTTC
421 CAAGCGAGCC CTGGGTTTGC GCAGGGAGCG GGCTGCTCTG GCGGTGGTTC CCGGAAACGC
481 AGCGCGCGCG ACCCTGGGTC TCGCACATTC TTCACGTCCG TTCCAGCGT CACCGGATC
541 TTCGCGGCTA CCCTTGTGGG CCCCCCGGCG ACGCTTCCTC GTCCGCCCTT AAGTCGGGAA
601 GGTTCCTTGC GGTTCGCGGG GTGCGGAGCG TGACAAACGG AAGCCGCACG TCTCACTAGT
661 ACCCTGCGAG ACGGACAGCG CCAGGGAGCA ATGGCAGCGC GCCGACCGCG ATGGGCTGTG
721 GCCAATAGCG GCTGCTCAGC AGGGCGCGCC GAGAGCAGCG GCCGGGAAGG GCGGCTGCGG
781 GAGGGGGGGT GTGGGGCGGT AGTGTGGGCC CTGTTCCCTG CCGCGCGGTT TTCCGCATTC
841 TGCAAGCCTC CGGAGCGCAC GTGGGCAGTC GGCTCCCTCG TTGACCGAAT CACCGACCTC
901 TCTCCCAGG ATCCACTAGT ATTAATCGT ACGCCTAGTA TTTAAATCGT ACGCCTAGTA
961 ACGGCGGCCA GTGTGCTGCA GATATCAAAG TCGACGGTAC CGGAGAGACC ATGCAGAGGT
1021 CGCCTCTGGA AAAGGCCAGC GTTGTCTCCA AACTTTTTTT CAGCTGGACC AGACCAATTT
1081 TGAGGAAAGG ATACAGACAG CGCCTGGAAT TGTACAGCAT ATACCAAATC CCTTCTGTG
1141 ATTCTGCTGA CAATCTATC CTAAAATTGG AAAGAGAATG GGATAGAGAG CTGGCTTCAA
1201 AGAAAAATCC TAAACTCATT AATGCCCTTC GCGGATGTTT TTTCTGGAGA TTTATGTTCT
1261 ATGGAATCTT TTATATTTTA GGGGAAGTCA CCAAAGCAGT ACAGCCTCTC TTACTGGGAA
1321 GAATCATAGC TTCTATGAC CCGGATAACA AGGAGGAACG CTCTATCGCG ATTTATCTAG
1381 GCATAGGCTT ATGCCCTTCT TTTATTGTGA GGACACTGCT CCTACACCCA GCCATTTTGT
1441 GCCTTCATCA CATTGGAATG CAGATGAGAA TAGCTATGTT TAGTTTGATT TATAAGAAGA
1501 CTTTAAAGCT GTCAAGCCGT GTTCTAGATA AAATAAGTAT TGGACAACCT GTTAGTCTCC
1561 TTTCCAACAA CCTGAACAAA TTTGATGAAG CACTTGCAAT GGCACATTTT GTGTGGATCG
1621 CTCTCTTTCG AGTGGCACTC CTCATGGGCG TAATCTGGGA GTTGTTCACG GCGTCTGCCT
1681 TCTGTGACT TGGTTTCCTG ATAGTCCCTG CCCTTTTTCG GGTGGGCTA GGGAGAATGA
1741 TGATGAAGTA CAGAGATCAG AGAGCTGGGA AGATCAGTGA AAGACTTGTG ATTACCTCAG
1801 AAATGATTGA AAACATCCAA TCTGTTAAGG CATACTGCTG GGAAGAAGCA ATGGAAAAAA

```

## Nucleotide Sequence Analysis (cont.)

```

1861 TGATTGAAAA CTTAAGACAA ACAGAACTGA AACTGACTCG GAAGGCAGCC TATGTGAGAT
1921 ACTTCAATAG CTCAGCCTTC TTCTTCTCAG GGTTCCTTGT GGTGTTTFTA TCTGTGCTTC
1981 CCTATGCACT AATCAAAGGA ATCATCCTCC GGAAATATT CACCAOCATC TCATTCTGCA
2041 TTGTTCTGCG CATGGCGGTC ACTOGGCAAT TTCCCTGGGC TGTACAAACA TGGTATGACT
2101 CTCTTGAGGC AATAAACAAA ATACAGGATT TCTTACAAA GCAAGBAATAT AAGACATTGG
2161 AATATAACTT AACGACTACA GAAGTAGTGA TGGAGAATGT AACAGCCTTC TGGGAGGAGG
2221 GATTTCGGGA ATTATTTGAG AAAGCAAAAC AAAACAATAA CAATAGAAAA ACTTCTAATG
2281 GTGATGACAG CCTCTTCTTC AGTAATTTCT CACTTCTTGG TACTCCTGTC CTGAAAGATA
2341 TTAATTTCAA GATAGAAAGA GGACAGTGTG TGGCGGTTGC TGGATCCACT GGAGCAGGCA
2401 AGACTTCACT TCTAATGATG ATTATGGGAG AACTGGAGCC TTCAGAGGGT AAAATTAAGC
2461 ACAGTGAAG AATTTTCATC TGTTCTCAGT TTCTCTGGAT TATGCTGGC ACCATTAAAG
2521 AAAATATCAT CTTTGGTGT TCTATGATG AATATAGATA CAGAAGCGTC ATCAAAGCAT
2581 GCCAACTAGA AGAGGACATC TCCAGTGTG CAGAGAAAGA CAATATAGTT CTGTGAGAGG
2641 GTGGAAATCAC ACTGAGTGA GGTCAACGAG CAAGAATTC TTTAGCAAGA GCAGTATACA
2701 AAGATGCTGA TTTGTATTTA TTAGACTCTC CTTTGGATA CCTAGATGTT TTAACAGAAA
2761 AAGAAATATT TGAAAGCTGT GTCTGTAAAC TGATGGCTAA CAAAAGTAGG ATTTTGGTCA
2821 CTTCTAAAAT GGAACATTTA AAGAAAGCTG ACAAATATT AATTTTGCAT GAAGGTAGCA
2881 GCTATTTTTA TGGGACATTT TCAGAACTCC AAAATCTACA GCCAGACTTT AGCTCAAAAC
2941 TCATGGGATG TGATTCCTTC GACCAATTTA GTGCAGAAAG AAGAAATTCA ATCCTAAGT
3001 AGACCTTACA CGTCTTCTCA TTAGAAGGAG ATGCTCCTGT CTCTGGACA GAAACAAAAA
3061 AACAATCTTT TAAACAGACT GGAGAGTTTG GGGAAAAAG GAAGAATTCT ATTCTCAATC
3121 CAATCAACTC TATACGAAAA TTTTCCATTG TGCAAAAGAC TCCCTTACAA ATGAATGGCA
3181 TCGAAGAGGA TTCTGATGAG CCTTAGAGA GAAGGCTGTC CTTAGTACCA GATTCTGAGC
3241 AGGGAGAGGC GATACTGCCT CGCATCAGCG TGATCAGCAC TGGCCCCACG CTTGAGCAC
3301 GAAAGAGGCA GTCTGTCTG AACCAGATGA CACACTCAGT TAACCAAGGT CAGAACATTC
3361 ACGGAAAGAC AACAGCATCC ACACGAAAAG TGTCACTGGC CCTCAGGCA AACTTGACTG
3421 AACTGATAT ATATTCAAGA AGGTATCTC AAGAACTGG CTTGGAATA AGTGAAGAAA
3481 TTAACGAAGA AGACTTAAAG GAGTGCCTTT TTGATGATAT GGAGAGCATA CCAGCAGTGA
3541 TACATGGAA CACATACCTT CGATATATTA CTGTCCCAA GAGCTTAATT TTTGTGTAA
3601 TTTGGTGCTT AGTAATTTTT CTGGCAGAGG TGGCTGCTTC TTTGGTTGTG CTGTGGCTCC
3661 TTGGAACAC TCCTCTTCAA GACAAAGGGA ATAGTACTCA TAGTAGAAT AACAGCTATG
3721 CAGTGATTAT CACCAGCACC AGTGTGATT ATGTGTTTTA CATTTACGTG GGAGTAGCCG
3781 ACCTTTGCT TGCTATGGA TTCTTCAGAG GTCTACCACT GGTGCATACT CTAATCAGAG
3841 TGTCGAAAT TTTACACCAC AAAATGTTAC ATTCTGTTCT TCAAGCACCT ATGTCAACCC
3901 TCAACACGTT GAAAGCAGGT GGGATTCTTA ATAGATTCTC CAAAGATATA GCAATTTTGG
3961 ATGACCTTCT GCCTCTTACC ATATTGACT TCATCCAGTT GTTATTAATT GTGATTGGAG
4021 CTATAGCAGT TGTGCGAGTT TTACAACCT ACATCTTTGT TGCAACAGTG CCAGTGATAG
4081 TGGCTTTTAT TATGTTGAGA GCATATTTCC TCCAAACCTC ACAGCAACTC AAACAAGTGG
4141 AATCTGAAGG CAGGAGTCCA ATTTTCACTC ATCTTGTAC AAGCTTAAAA GGACTATGGA
4201 CACTTCGTGC CTTCCGACGG CAGCCTTACT TTGAAACTCT GTTCCACAAA GCTCTGAATT
4261 TACATACTGC CAACTGGTTC TTGTACCTGT CAACACTGCG CTGGTTCCAA ATGAGAATAG
4321 AAATGATTTT TGTCACTTTC TTCAATGCTG TTACCTTCAT TTCCATTTTA ACAACAGGAG
4381 AAGGAGAAGG AAGAGTTGGT ATTATCCTGA CTTTAGCCAT GAATATCATG AGTACATTGC
4441 AGTGGGCTGT AAACCTCAGC ATAGATGTGG ATAGCTTGAT GCGATCTGTG AGCCGAGTCT
4501 TTAAGTTTAT TGACATGCCA ACAGAAGGTA AACCTACCAA GTCAACCAA CCATACAAGA
4561 ATGGCCAACT CTCGAAAGTT ATGATTATTG ACAAATCACA CGTGAAGAAA GATGACATCT
4621 GGCCCTCAGG GGGCCAAATG ACTGTCAAAG ATCTCACAGC AAAATACACA GAAGGTGGA
4681 ATGCCATATT AGAGAACATT TCCTTCTCAA TAAGTCTTGG CCAGAGGGTG GGCCTCTTGG
4741 GAAGAAGTGG ATCAGGGAAG AGTACTTTGT TATCAGCTTT TTTGAGACTA CTGAACACTG
4801 AAGGAGAAAT CCAGATCGAT GGTGTGCTCT GGGATTCAT AACTTTGCAA CAGTGGAGGA
4861 AAGCCTTTGG AGTGATACCA CAGAAAGTAT TTATTTTTTC TGGAAATTT AGAAAAAAGT
4921 TGGATCCCTA TGAACAGTGG AGTGATCAAG AAATATGGAA AGTTGCAGAT GAGGTTGGCC
4981 TCAGATCTGT GATAGAACAG TTTCCCTGGG AGCTTGACTT GTCTCTGTG GATGGGGGCT
5041 GTGTCTTAAG CCATGGCCAC AAGCAGTTGA TGTGCTTGGC TAGATCTGTT CTCAGTAAGG
5101 CGAAGATCTT GCTGCTTGAT GAACCCAGTG CTCATTTGGA TCCAGTAACA TACCAATAA
5161 TTAGAAGAAC TCTAAACAA GCATTTGCTC ATTGCACAGT AATTCTCTGT GAACACAGGA
5221 TAGAAGCAAT GCTGGAATGC CAACAATTTT TGGTCATAGA AGAGAACAAA GTGCGGCACT

```



## Nucleotide Sequence Analysis (cont.)

```

5281 ACGATTCCAT CCAGAACTG CTGAACGAGA GGAGCCTCTT CCGGCAAGCC ATCAGCCCCCT
5341 CCGACAGGGT GAAGCTCTTT CCCCACCGGA ACTCAAGCAA GTGCAAGTCT AAGCCCCAGA
5401 TTGCTGCTCT GAAAGAGGAG ACAGAAGAAG AGGTGCAAGA TACAAGGCTT TAGAGAGCAG
5461 CATAAATGTT GACATGGGAC ATTTGCTCAT GGAATTGGAG AAATCGTACG CTTAGGACGC
5521 GTAATAAAAT GAGGAAATG CATCGCATTC TCTGACCGGT TACGCGGGA GGTGCTGAGG
5581 TAGGATGAGA CCGCACCAG GTGCAGACCC TGGAGTGTG GCGTAAACA TATTAGGAAC
5641 CAGCCTGTGA TGCTGGATGT GACCGAGGAG CTGAGGCCCG ATCACTGGT GCTGGCCTGC
5701 ACCCGCGCTG AGTTTGGCTC TAGCGATGAA GATACAGATT GAGGTACTGA AATGTGTGGG
5761 CGTGGCTTAA GGGTGGGAAA GAATATATAA GGTGGGGGTC TCATGTAGTT TTGTATCTGT
5821 TTTGCAGCAG CCGCGCCAT GAGCGCCAAC TCGTTTGATG GAAGCATTGT GAGCTCATAT
5881 TAGCAACCG GCATGCCCC ATGCGCGGG GTGCGTCAGA ATGTGATGG CTCCAGCATT
5941 GATGGTGGCC CGTCCCTGCC CGCAAACTCT ACTACCTTGA CCTACGAGAC CGTGTCTGGA
6001 ACGCGGTTGG AGACTGCAGC CTCGCGCGCC GCTTCAGCCG CTGCAGCCAC CGCCGCGCGG
6061 ATGTGACTG ACTTTGCTTT CCTGAGCCCG CTTCGAAGCA GTGCAGCTTC CCGTTCATCC
6121 GCCCGCGATG ACAAGTTGAC GGTCTTTTTC GCACAATTGG ATTCTTTGAC CCGGGAATTT
6181 AATGTGCTTT CTCAGCAGCT GTTGGATCTG CGCCAGCAGG TTCTGCGCT GAAGGCTTCC
6241 TCCCCCTCCA ATGCGGTTTA AAACATAAAT AAAAACCAGA CTCTGTTTGG ATTTTGATCA
6301 AGCAAGTGTG TTGCTGTCTT TATTTAGCGG TTTTGGCGC GCGGTAGGCC CCGGACCAGC
6361 GGTCTCGGTC GTTGAGGGTC CTGTGTATTT TTTCAGGAC GTGGTAAAGG TGACTCTGGA
6421 TGTTCAAGATA CATGGGCATA AGCCCGTCT TGGGGTGGAG GTAGCACCAC TGCAGAGCTT
6481 CATGCTGCGG GGTGCTGTG TAGATGATCC AGTCGTAGCA GGAGCGCTGG GCGTGGTCCC
6541 TAAAAATGTC TTTCAGTAGC AAGCTGATTG CCAGGGGCG GCGCTTGGTG TAAGTGTFTA
6601 CAAAGCGGTT AAGCTGGGAT GGTGTCATAC GTGGGATAT GAGATGCATC TTGGACTGTA
6661 TTTTLAGGTT GGTATGTTT CCAGCCATAT CCTCCGGGG ATTCATGTTG TGCAGAACCA
6721 CCAGCACAGT GTATCCGGTG CACTTGGGAA ATTTGTATG TAGCTTAGAA GGAATGCGT
6781 GGAAGAACTT GGAGACGCCC TTGTGACCTC CGAGATTTTC CATGCATTGG TCCATAATGA
6841 TGGCAATGGG CCCACGGGCG GCGGCCTGGG CGAAGATATT TCTGGGATCA CTAACCTCAT
6901 AGTTGTGTTT CAGGATGAGA TCGTCATAGG CCATTTTAC AAAGCGGGG CCGAGGGTGC
6961 CATCTGCGG TATAATGTTT CCATCCGGCC CAGGGGCGTA GTTACCCTCA CAGATTGCA
7021 TTTCCACGCG TTGAGTTTCA GATGGGGGA TCATGTCTAC CTGCGGGCG ATGAAGAAAA
7081 CCGTTTCCGG GGTAGGGGAG ATCAGCTGGG AAGAAAGCAG GTTCCCTGAGC AGCTCCGACT
7141 TACCGCAGCC GGTGGGCCCC TAAATCACAC CTATTACCG CTGCAACTGG TAGTTAAGAG
7201 AGCTGCAGCT GCCGTCTATC CTGAGCAGGG GGGCCACTTC GTTAAAGCATG TCCCTGACTT
7261 GCATGTTTTT CCTGACCAAA TGCGCCACAA GCGCTCGCC CCCCAGCGAT AGCAGTCTT
7321 GCAAGGAAGC AAGTTTTTTC AACGGTTTGA GCGCGTCCGC CGTAGGCATG CTTTTGAGCG
7381 TTTGACCAAG CAGTTCCAGG CCGTCCACAC GCTCGGTAC GTGCTCTACG GCATCTCGAT
7441 CCAGCATATC TCCTCGTTTC GCGGGTTGGG GCGGCTTTCG CTGTACGGCA GTAGTGGGTG
7501 CTGTTCCAGA CCGGCCAGGG TCATGTCMTT CCACGGGCGC AGGGTCTCTG TCACCGTAGT
7561 CTGGGTACCG GTGAAGGGT GCGCTCCGG CTGCGCGCTG GCCAGGGTGC GCTTGAGGCT
7621 GGTCTGCTG GTGCTGAAGC GCTGCCGCTC TTGCGCCTGC GCGTCCGCCA GGTAGCATTT
7681 GACCATGGTG TCATAGTCCA GCGCTCCGC GCGTGGCCC TTGGCGCGCA GCTTGCCCTT
7741 GGAGGAGGCG CCGCACGAGG GGCAGTGCAG ACTTTTAAG GCGTAGAGCT TGGGCGCGAG
7801 AAATACCGAT TCCGGGAGT AGGCATCCGC GCGCAGGCC CCGCAGACGG TCTCCCATTC
7861 CACGAGCCAG GTGAGCTCTG CCGGTTCCGG GTCAAAAACC AGGTTTCCCC CATGCTTTTT
7921 GATGCGTTTC TTACCTCTGG TTTCATGAG CCGGTGTCCA CGCTCGTGA CGAAAAGGCT
7981 GTCCTGTGCC CCGTATACAG ACTTGAGAGG CCTGTCTCTG AGCGGTGTTT CCGGTCCTC
8041 CTCGTATAGA AACTCGGACC ACTCTGAGAC GAAGGCTCGC GTCCAGGCCA GCACGAAGGA
8101 GGCTAAGTGG GAGGGGTAGC GGTCTGTGTC CACTAGGGGG TCCACTCGCT CCAGGGTGTG
8161 AAGACACATG TCGCCCTCTT CCGCATCAAG GAAGGTGATT GGTATTATAGG TGTAGGCCAC
8221 GTGACCGGGT GTTCTTGAAG GGGGGCTATA AAAGGGGGTG GGGGCGGCTT CGTCTCACT
8281 CTCTTCCGCA TCGCTGTCTG CGAGGGCCAG CTGTTGGGCT GAGTACTCCC TCTCAAAAGC
8341 GGGCATGACT TCTGCGCTAA GATTGTGAGT TTCCAAAAC GAGGAGGATT TGATATTAC
8401 CTGCCCCGCG GTGATGCCCT TGAGGGTGGC CCGCTCCATC TGGTCAGAAA AGACAATCTT
8461 TTTGTGTGTA AGCTTGGTGG CAAACGACCC GTAGAGGGCG TTGGACAGCA ACTTGGCGAT
8521 GGAGCGCAGG GTTTCGTTTT TGTCCGATC GCGCGCTCC TTGGCCGCGA TGTTAGCTG
8581 CAGCTATTCC CGGCAACGCG ACCGCCATTG GGAAGACG GTGGTGCCTG CGTCCGGCAC
8641 CAGGTGCACG CGCCAACGCG GGTGTGTCAG GGTGACAAGG TCAACGCTGG TGGCTACCTC

```

## Nucleotide Sequence Analysis (cont.)

```

8701 TCCGCGTAGG CGCTCGTTGG TCCAGCAGAG GCGGCGGCC TTGCGCGAAC AGAATGGCGG
8761 TAGTGGGTCT AGCTGCGTCT CGTCGGGGG GTCTGCGTCC ACGGTAAAGA CCCC GGCCAG
8821 CAGGCGCGCG TOGAAGTAGT CTATCTTGCA TCCTTGCAAG TCTAGCGCCT GCTGCCATGC
8881 GCGGCGGCGA AGCGCGCGCT CGTATGGGTT GAGTGGGGGA CCCCATGGCA TGGGTTGGGT
8941 GAGCGCGGAG GGGTACATGC CGCAAAATGTC GTAAACGTAG AGGGGCTCTC TGAGTATTCC
9001 AAGATATGTA GGGTAGCATC TTCCACCGCG GATGCTGGCG CGCACGTAAT CGTATAGTTC
9061 GTGCGAGGGA GCGAGGAGGT CCGGACCGAG GTTGCTACGG GCGGGCTGCT CTGCTCGGAA
9121 GACTATCTGC CTGAAGATGG CATGTGAGTT GGATGATATG GTTGACGCT GGAAGACGTT
9181 GAGCTGGCG TCTGTGAGAC CTACCGCGTC ACGCACGAAG GAGGCGTAGG AGTCGCGCAG
9241 CTTGTTGACC AGCTCGGCGG TGACCTGCAC GTCTAGGGCG CAGTAGTCCA GGGTTTCTTT
9301 GATGATGTCA TACTTATCCT GTCCCTTTT TTTCCACAGC TCGCGGTTGA GGACAACTC
9361 TTGCGGCTCT TTCCAGTACT CTTGGATCGG AAACCCGTCG GCCTCCGAAC GGTAAAGAGCC
9421 TAGCATGTAG AACTGGTTGA CGGCCTGGTA GCGGCAGCAT CCTTTTCTA CCGGTAGCGC
9481 GTATGCTGCG GCGGCCTTCC GGAGCGAGGT GTGGGTGAGC GCAAGGTTGT CCTAACCAT
9541 GACTTTGAGG TACTGGTATT TGAAGTCAGT GTGTCGCGAT CCGCCCTGCT CCCAGAGCAA
9601 AAGTCCGCTG CGCTTTTGG AACGCGGGTT TGGCAGGGCG AAGGTGACAT CGTTGAAAAG
9661 TATCTTTCCC GCGCGAGGCA TAAAGTTGCG TGTGATGCGG AAGGGTCCCG GCACCTCGGA
9721 ACGGTTGTTA ATTACCTGGG CCGCGAGCAC GATCTGCTCG AAGCCGTGTA TGTGTGGGCC
9781 CACGATGTAA AGTTCCAAGA AGCGCGGGGT GCGCTGTATG GAGGCAATT TTTTAAGTTC
9841 CTGCTAGGTG AGCTCCTCAG GGGAGCTGAG CCGGTGTTCT GACAGGGCCC AGTCTGCAAG
9901 ATGAGGGTTG GAAGCGAAGA ATGAGCTCCA CAGGTCAAGG GCCATTAGCA TTTGCGAGTG
9961 GTCCGAAAG GTCCCTAACT GCGGACCTAT GGCCATTTT TCTGGGTTGA TGCAGTAGAA
10021 GGTAAAGCGG TCTTGTTCCT AGCGGTCCCA TCCAAGTCC ACGGCTAGGT CTGCGCGCGC
10081 GGTCAACAGA GGCTCATCTC CGCGAACTT CATAACGAGC ATGAAGGGCA CGAGCTGCTT
10141 CCCAAAGGCC CCCATCCAAG TATAGGTCTC TACATCGTAG GTGACAAAGA GACGCTCGGT
10201 GCGAGGATGC GAGCGATCG GGAAGAACTG GATCTCCGCG CACCAAGTTG AGGAGTGGCT
10261 GTTGATGTGT TGAAAGTAGA AGTCCCTGCG ACGGGCGGAA CACTCGTCTC GGCTTTTGTA
10321 AAAACGTGCG CAGTACTGGC AGCGGTGCAC GGGCTGTACA TCCTGCACGA GGTGACCTG
10381 ACGACCGCGC ACAAGGAAGC AGAGTGGGAA TTTGAGCCCC TCGCCTGGCG GGTGTCCTG
10441 GTGCTCTTCT ACTTCGGCTG CTTGTCTTTC ACCGTCTGGC TGCTCGAGGG GAGTTATGGT
10501 GGATCGGACC ACCACGCGCG GCGAGCCCAA AGTCCAGATG TCCGCGCGCG GCGGTGCGAG
10561 CTTGATGACA ACATCGCGCA GATGGGAGCT GTCCATGCTC TGGAGCTCCC GCGGCGACAG
10621 GTCAGGCGGG AGCTCCTGCA GGTTTACCTC GCATAGCGCG GTGAGGCGCG GCGGCTAGCTC
10681 CAGGTGATAC CTGATTTCCA GGGGCTGGTT GGTGGCGCGC TCGATGACTT GCAAGAGGCC
10741 GCATCCCGCG GCGCGACTA CGGTACCGCG CGGCGGCGCG TGGGCTGCTT GGGTGTCTTT
10801 GGATGATGCA TCTAAAAGCG GTGACGCGGG CCGGCCCCCG GAGGTAGGGG GGGCTCGGGA
10861 CCCGCGGGGA GAGGGGGCAG GGGCACGTCG GCGCGCGCGG CCGGCAGGAG CTGGTCTGTC
10921 GCGCGGAGGT TGCTGGCGAA CCGGACGACG CCGCGGTGTA TCTCTGAAT CTGGCGCTTC
10981 TCGGTGAAGA CGACGGGCCC GGTGAGCTTG AACCTGAAAG AGAGTTGAC AGAATCAATT
11041 TCGGTGTCGT TGACGGCGGC CTGGCGCAAA ATCTCCTGCA CGTCTCTGA GTTGCTTGA
11101 TAGGCGATTT CCGCCATGAA CTGCTGATC TCTTCTCTT GGAGATCTCC GCGTCCGGCT
11161 CGCTCCACGG TGGCGGCGAG GTGTTGGAG ATGCGGGCCA TGAGCTCGCA GAAGGCGTTG
11221 AGGCTCTCCT CGTTCCAGAC GCGGCTGTAG ACCACGCCCC CTTCGGCATC GCGGCGCGC
11281 ATGACCACCT GCGCGAGATT GAGCTCCACG TCGCGGCGCA AGACGGCGTA GTTTCGCAAG
11341 CGCTGAAAAG GGTAGTTGAG GGTGTTGGCG GTGTGTTCTC CCACGAAGAA GTACATAACC
11401 CAGCGTGCAG ACGTGGATTG GTTGATATCC CCCAAGGCTC CAAGGCGCTC CATGGCTCG
11461 TAGAAGTCCA CCGCGAAGTT GAAAAACTGG GAGTTGCGCG CCGACACGGT TAACCTCTCC
11521 TCCAGAAAGC GGTAGAGCTC GCGGACAGTG TCGCGCACCT CCGGCTCAA GGTACAGGG
11581 GCCTCTTCTT CTTCAATCTC CTCCTCCATA AGGCGCTCCC CTCTCTCTC TTCTTCTGGC
11641 GCGGCTGGGG GAGGGGGGAC ACGGCGGCGA CGACGGCGCA CCGGAGGCG GTCCGACAAAG
11701 CGCTCGATCA TCTCCCGCG GCGACGCGCG ATGGTCTCGG TGACGCGCGG GCCGTCTCTG
11761 CCGGGGCGCA GTTGGAAGAC GCCGCGCGTC ATGTCCGCT TATGGGTTGG CCGGGGGCTG
11821 CCGTCCGGCA GGGATACGGC GCTAACGATG CATCTCAACA ATTGTTGTGT AGGTACTCG
11881 CCACCGAGCG ACCTGAGCGA GTCCGATCG ACCGGATCG AAAACCTCTC GAGAAAGCGG
11941 TCTAACCAGT CACAGTCGCA AGGTAGGCTG ACCACGCTG CCGGCGGCGG CCGGTGGGGG
12001 TCGGGGTTGT TCTGCGCGA GGTCTGCTG ATGATGTAAT TAAAGTAGGC GGTCTTGAGA
12061 CCGCGGATGG TCGACAGAAG CACCATGTCC TTGGGTCCGG CCTGCTGAAT GCGCAGCGCG

```

## Nucleotide Sequence Analysis (cont.)

```

12121 TCGGCCATGC CCCAGGCTTC GTTTTGACAT CGGCGCAGGT CTTTGTAGTA GTCTTGCATG
12181 AGCCTTTCTA CCGGCACTTC TTCTTCTCCT TCCTCTGTGC CTGCATCTCT TGCACTCTATC
12241 GCTACGGGGG CGGCGGAGTT TGGCGGTAGG TGGGCGCCCTC TTCTTCCCAT GCCTGTGACC
12301 CCGAAGCCCC TCATCGGCTG AAGCAGGGCC AGGTCGGCGA CAACGGGCTC GGCTAATATG
12361 GCCTGCTGCA CCTGCGTGAG GGTAGACTGG AAGTCATCCA TGTCCACAAA GCGGTGGTAT
12421 GCGCCCGTGT TGATGGTGTA AGTGCAGTTG GCCATAAOCG ACCAGTTAAC GGTCTGGTGA
12481 CCGGCTGTGG AGAGCTCGGT GTACCTGAGA CGCGAGTAAG CCCTTGAGTC AAAGACGTAG
12541 TCGTTGCAAG TCGGCACCAG GTACTGATAT CCCACCAAAA AGTGCGGCGG CCGCTGGCGG
12601 TAGAGGOGCC AGCGTAGGGT GCGCGGGGCT CCGGCGGCGA GGTCTTCCAA CATAGGCGA
12661 TGATATCCGT AGATGTACCT GGACATCCAG GTGATGCCGG CCGCGGTGGT GGAGCGCGC
12721 GGAAGTCCG CGACCGGGTT CCAGATGTTG CGCAGCGCCA AAAAGTGCTC CATGGTCCGG
12781 ACCTCTTGGC CGGTGAGGGG TCGCGAGTCG TTGACGCTCT AGACCGTGCA AAAGAGAGAC
12841 CTGTAAGCGG GCACTCTTCC GTGGTCTGGT GGATAAATTC GCACGGGTAT CATGGCGGAC
12901 GACCGGGGTT CGAAGCCCGG ATCGGGCGGT CCGCGGTGAT CCATGCGGTT ACCGCCGGCG
12961 TGTGAAACCC AGGTGTGCGA CGTCAGACAA CCGGGGAGGG CTCCTTTTGG CTTCCTTCCA
13021 GCGCGGGCGG CTGCTGCGCT AGCTTTTTTG GCCACTGGCC GCGCGCGGGG TAAGCGGTTA
13081 GCGTGGAAG CGAAGCATT AAGTGGCTCG CTCCTGTAG CCGGAGGCTT ATTTTCCAAG
13141 GGTGTAGTGG CAGGACCCCG GGTTCGAGTC TCGGCGGGGC CGGACTGCGG CGAAGCGGGG
13201 TTTGCCCTCC CGTCATGCAA GACCCCGCTT GCAAAATCCT CCGGAAACAG GGACGAGCCC
13261 CTTTTTTGCT TTTCCAGAT GCATCGGGTG CTGCGGCAGA TCGCCCCCCC TCCTCAGCAG
13321 CGCAAGAGC AAGAGCAGCG GCAGACATCC AGGGCACCCT CCCTTCTTCC TACCGCGTCA
13381 GAGGGGGCAA CATCCGCGGC TGACCGGGCG GCAGATGGTG ATTACGAACC CCGCGGGCGC
13441 GGGGCCGGGC ACTACCTGGA CTGTGAGGAG GCGGAGGGCC TGGCGGGGCT AGGAGCGCCC
13501 TCTCTGAGC GACACCCAAG GGTGCAGCTG AAGCGTGACA CCGCGGAGGC GTACGTGCCG
13561 CCGCAGAAC TGTTCGCGA CCGCGAGGGA GAGGAGCCCG AGGAGATGCG GGATCGAAG
13621 TTCCACGCG GCGCGAGTT GCGGCATGGC CTGAACCGCG AGCGGTTGCT GCGCGAGGAG
13681 GACTTTGAGC CCGAGCGGCG GACCGGGATT AGTCCCGGCG GCGCACAGCT GCGCGCGGCC
13741 GACCTGGTAA CCGCGTACGA GCAGACGGTG AACCAGGAGA TTAAGTTTCA AAAAGCTTT
13801 AACCAACCAG TCGCAGCGCT TGTGGCGGCG GAGGAGGTGG CTATAGGACT GATGCATCTG
13861 TCGGACTTTG TAAGCGCGCT GGAGCAAAAC CCAATAGCA AGCCGCTCAT GGCGCAGCTG
13921 TTCTTTATAG TGCAGCACAG CAGGACAAC GAGGCATTCA GGGATGCGCT GCTAAACATA
13981 GTAGAGCCCG AGGGCGCTG GCTGCTCGAT TTGATAAACA TTCTGCAGAG CATAGTGGTG
14041 CAGGAGGCGA GCTTGAGCCT GGCTGACAAG GTGGCCGCCA TTAAGTTTCA CATGCTCAGT
14101 CTGGGCAAGT TTTACGCCCG CAAGATATAC CATACCCCTT ACGTTCCCAT AGACAAGGAG
14161 GTAAAGATCG AGGGTTCTA CATGCGCATG GCGTTGAAGG TGCTTACCTT GAGCGACGAC
14221 CTGGGCGTTT ATCGCAACGA GCGCATCCAC AAGGCCGTGA GCGTGAGCCG GCGGCGCGAG
14281 CTCAGCGACC GCGAGCTGAT GCACAGCCTG CAAAGGGCCC TGGCTGGCAC GGGCAGCGGC
14341 GATAGAGAGG CCGAGTCTTA CTTTGACGCG GCGCTGAGC TGGCTGGGC CCCAAGCCGA
14401 CCGCCCTTGG AGGCAGCTGG GCGCGGACCT GGGCTGGGG TGGCACCCGC GCGCGCTGGC
14461 AACGTCCGGC GCGTGAGGA ATATGACGAG GACGATGAGT ACGAGCCAGA GGACCGCGAG
14521 TACTAAGCGG TGATGTTTCT GATCAGATGA TGCAAGACGC AACGAGCCCG GCGGTGCGGG
14581 CCGCGCTGCA GAGCCAGCCG TCCGGCCTTA ACTCCACGGA CGACTGGCGC CAGGTCTATG
14641 ACCGCATCAT GTGCTGACT GCGCGTAACC CTGACGCGTT CCGGCAGCAG CCGCAGGCCA
14701 ACCGGCTCTC CGCAATTCTG GAAGCGGTGG TCCCGGCGCG CGCAAAACCC ACGCAGAGA
14761 AGGTGCTGGC GATCGTAAAC GCGCTGGCG AAAACAGGGC CATCCGGCCC GATGAGGCGG
14821 GCTTGGCTA CGACGCGCTG CTTGAGCGCG TGGCTCGTTA CAACAGCGGC AACGTGCAGA
14881 CCAACCTGGA CCGCTGGTG GGGGATGTGC GCGAGGCCGT GCGCAGCGT GAGCGCGCGC
14941 AGCAGCAGGG CAACCTGGGC TCCATGTTG CACTAAACGC CTTCTGAGT ACACAGCCCG
15001 CCAACGTGCC GCGGGGACAG GAGGACTACA CCAACTTTGT GAGCGCACTG CCGCTAATGG
15061 TGACTGAGAC ACCGCAAGT GAGGTGTACC AGTCCGGGCC AGACTATTTT TTCCAGACCA
15121 GTAGACAAGG CCGCAGACC GTAAACCTGA GCCAGGCTTT CAAGAACTTG CAGGGGCTGT
15181 GGGGGGTGCG GCTCCACA GCGCAGCGCG CGACCGTGTG TAGCTTGCTG ACGCCCAACT
15241 CCGCCCTGTT GCTGCTGCTA ATAGCGCCCT TCACGACAG TGGCAGGGTG TCCGGGACA
15301 CATACCTAGG TCACCTGCTG ACACGTATCC GCGAGGCCAT AGGTGAGGCG CATGTGACG
15361 AGCATACTTT CCAGGAGATT ACAAGTGTCA GCGCGCGCT GGGGCGAGG GACACGGGCA
15421 GCCTGGAGGC AACCTGAAC TACCTGCTGA CCAACCGGGC GCAGAAGATC CCCTCGTTGC
15481 ACAGTTTAAA CAGCGAGGAG GAGCGCATCT TCGCTATGTT GCAGCAGAGC GTGAGCCTTA

```

## Nucleotide Sequence Analysis (cont.)

```

15541 ACCTGATGCG CGACGGGGTA ACGCCAGCG TGGCGCTGGA CATGACCGCG CGCAACATGG
15601 AACCGGGCAT GTATGCCTCA AACCGGCGT TTATCAATCG CCTAATGGAC TACTTGCATC
15661 GCGCGGCGCG CGTGAACCCC GAGTATTTCA CCAATGCCAT CTGAACCGG CACTGGCTAC
15721 CGCCCCCTGG TTTCTACACC GGGGATTTG AGGTGCCCGA GGGTAACGAT GATTCCTCT
15781 GGGACGACAT AGACGACAGC GTGTTTTCCC CGCAACCGCA GACCTGCTA GAGTTGCAAC
15841 AGCGGAGACA GGCAGAGGCG GCGCTGCGAA AGGAAAGCTT CCGCAGGCCA AGCAGCTTGT
15901 CCGATCTAGG CCGTGCAGCC CCGCGGTCAG ATCGGAGTAG CCCATTTCCA AGCTTGATAG
15961 GGTCTTTTAC CAGCACTGCG ACCACCGGCC CGCGCTGCT GGGGAGGAG GAGTACCTAA
16021 ACAACTCGCT GCTGCAGCCG CAGCGCGAAA AGAACCCTGC TCCGCGATTT CCCAACAAAG
16081 GGTATGAGAG CCTAGTGGAC AAGATGAGTA GATGGAAGAC GTATGCGCAG GAGCACAGGG
16141 ATGTGCCCGG CCGCGGCGCG CCCACCGGTC GTCAAAGGCA CGACCGTCAG CGGGGTCTCG
16201 TGTGGGAGGA CGATGACTCG GCAGACGACA GCAGCGTCTT GGATTTGGGA GCGAGTGGCA
16261 ACCCGTTTGC GCACCTTGGC CCGAGGCTGG GGAGAATGTT TTAACAAAAA AAAAAAAAG
16321 CATGATGCAA AATAAAAAAC TCACCAAGGC CATGGCACCG AGCGTTGGTT TTCTTGTTAT
16381 CCCCTTAGTA TGCAGGCGCG GCGGATGTAT GAGGAAGGTC CTCTCCCTC CTACGAGAGC
16441 GTGGTGAGCG CGGCGCCAGT GGGGCGGGCG CTGGGTTCCC CCTTGGATGC TCCCTGGAC
16501 CCGCGGTTTG TGCCCTCGCG GTACCTGCGG CCTACCGGGG GGAGAAACAG CATCGTTTAC
16561 TCTGAGTTGG CACCCCTATT CGACACCGCC CGTGTGTACC TTGTGGACAA CAAGTCAACG
16621 GATGTGGCAT COCTGAACTA CCAGAACGAC CACAGCAACT TTCTAAOCAC GGTCAATTCA
16681 AACATGACT ACAGCCCGGG GGAGGCAAGC ACACAGACCA TCAATCTTGA CGACCGTTGC
16741 CACTGGGGCG GCGACCTGAA AACCATCTCG CATACCAACA TGCCAAATGT GAACGAGTTC
16801 ATGTTTACCA ATAAGTTTAA GCGCGGGGTC ATGGTGTGCG GCTCGCTTAC TAAGGACAA
16861 CAGGTGGAGC TGAAATATGA GTGGGTGGAG TTCACGCTGC CCGAGGGCAA CTACTCGAG
16921 ACCATGACCA TAGACCTTAT GAACAAAGCG ATCGTGGAGC ACTACTTGAA AGTGGGCGAG
16981 CAGAACGGGG TTCTGGAAAG CGACATCGGG GTAAAGTTTG ACACCGCAA CTTCAGACTG
17041 GGGTTTGACC CAGTCACTGG TCTTGTCATG CCTGGGGTAT ATACAAACGA AGCCTTCCAT
17101 CACGACATCA TTTTGCTGCC AGGATGCGGG GTGACTTCA CCCACAGCG CCTGAGCAAC
17161 TTGTTGGGCA TCCGCAAGCG GCAACCCTTC CAGGAGGGCT TTAGGATCAC CTACGATGAC
17221 CTGGAGGGTG GTAACATTCC CGCACTGTTG GATGTGGACG CCTACCAGGC AAGCTTAAAA
17281 GATGACACCG AACAGGGCGG GGATGGCGCA GCGCGCGGCA ACAACAGTGG CAGCGGCGCG
17341 GAAGAGAACT CCAACGCGCG AGCGCGGCGA ATGCAGCGCG TGGAGGACAT GAACGATCAT
17401 GCCATTTCGG CGCACACCTT TGCCACAGCG GGGAGGAGA AGCGCGCTGA GGCCGAGGCA
17461 GCGGCAGAAG CTGCCGCCCC CGCTGGGCAA CCCGAGGTGC AGAAGCCTCA GAAGAAACCG
17521 GTGATCAAAAC CCCTGACAGA GGACAGCAAG AAACGCAGTT ACAACCTAAT AAGCAATGAC
17581 AGCACCTTCA CCCAGTACCG CAGCTGGTAC CTTGCATACA ACTACGGCGA CCCTCAGACC
17641 GCGATCGGCT CATGGACCTT CCTTTGCACT CTTGACGTAA CCTGGGGCTC GGAGCAGGTC
17701 TACTGGTGTG TGCCAGACAT GATGCAAGAC CCGGTGACCT TCGGCTCCAC GAGCCAGATC
17761 AGCAACTTTC CGGTGGTGGG CGCCGAGCTG TTGCGCGTGC ACTCCAAGAG CTTCTACAAC
17821 GACCAGGCGG TCTACTCCCA GCTCATCGC CAGTTTACCT CTCTGACCCA CGTGTTCAT
17881 CGCTTTCCCG AGAACCAGAT TTTGGGCGCG CCGCCAGCGC CCACCATCAC CACCGTCAGT
17941 GAAAACGTTT CTGCTCTCAC AGATCACGGG ACGCTACCGC TGCGCAACAG CATCGGAGGA
18001 GTCCAGCGAG TGACCATTAC TGACGCCAGA CGCCGACCTT GCGGCTACCT TTACAAGGCC
18061 CTGGGCATAG TCTCGCGCGG CGTCTATCG AGCCGCACTT TTTGAGCAA CATGTCCATC
18121 CTTATATGCG CCAGCAATAA CACAGGCTGG GGCCTGCGCT TCCCAAGCAA GATGTTTGGC
18181 GCGGCAAGA AGCGCTCCGA CCAACACCCA GTGCGCGTGC GCGGGCACTA CCGCGCGCCC
18241 TGGGGCGGCG ACAACCGGG CGCACTGGG CGCACCAAGC TCGATGACCG CATTGACGCG
18301 GTGGTGGAGG AGGCGCGCAA CTACAGCGCC ACGCGGCCAC CAGTGTCCAC AGTGGAGCGG
18361 GCCATTGAGA CCGTGGTGCG OGGAGCCCGG CGTTATGCTA AAATGAAGAG ACGGCGGAGG
18421 CGCGTAGCAC GTCGCCACCG CCGCGGACCC GGCCTGCGG CCEAACGCGC GCGGCGGCGC
18481 CTGCTTAACC GCGCACGTGC CACCGGCGCA CCGCGGCGCA TGCGGGCGCG TCGAAGGCTG
18541 GCGGCGGGTA TTGTCACTGT GCCCCCGAGG TCCAGGCGAC GAGCGGCGCG CGCAGCAGCC
18601 GCGGCCATTA GTGCTATGAC TCAGGGTGGC AGGGGCAACG TGTAAGGGT GCGCGACTCG
18661 GTTAGCGGCC TGCGCGTGCC CGTGCGCAAC CGCCCCCGC GCAACTAGAT TGCAAGAAAA
18721 AACTACTTAG ACTCGTACTG TTGTATGTAT CCGCGGCGCG CCGCGCGCAA CGAAGCTATG
18781 TCCAAGCGCA AAATCAAAGA AGAGATGCTC CAGGTCAATC CGCGGAGAT CTATGCCCCC
18841 CCGAAGAAAG AAGAGCAGGA TTACAAGCCC CGAAAGCTAA AGCGGGTCAA AAAGAAAAAG
18901 AAAGATGATG ATGATGATGA ACTTGACGAC GAGGTGGAAC TGCTGCACCG AACCAGCCCC

```

## Nucleotide Sequence Analysis (cont.)

18961	AGGCGGCGGG	TACAGTGGAA	AGGTGGACGC	GTAAGACGTG	TTTTCGGACC	CGGCACCACC
19021	GTAGTTTTTA	CGCCCGGTGA	GCGCTCCACC	CGCACCTACA	AGCGCGTGTA	TGATGAGGTG
19081	TACGGCGAAG	AGGACCTGCT	TGAGCAGGGC	AACGAGCGCC	TGGGGGAGTT	TGCCTAOGGA
19141	AAGCGGCATA	AGGACATGTT	GGCGTTGCCG	CTGGACGAGG	GCAACCCAAC	ACTTAGCCTA
19201	AAGCCCGTGA	CACTGCAGCA	GGTGTGCCC	ACGCTTGAC	CGTCCGAAGA	AAAGCGGGGC
19261	CTAAAGCGCG	AGTCTGGTGA	CTTGGCACCC	ACCGTGCAGC	TGATGGTACC	CAAGCGCCAG
19321	CGACTGGAAG	ATGTCTTGA	AAAAATGACC	GTGGAGCCTG	GGCTGGAGCC	CGAGGTCCGC
19381	GTGCGGCCAA	TCAAGCAGGT	GGCACCGGGA	CTGGGCGTGC	AGACCGTGA	CGTTTACAGATA
19441	CCCACCACCA	GTAGCACTAG	TATTGCCACT	GCCACAGAGG	GCATGGAGAC	ACAAACGTCC
19501	CGGTTGCCT	CGGCGGTGGC	AGATGCCCGG	GTGCAGGGGG	CGCTGGGGC	CGCGTCCAAA
19561	ACCTCTACGG	AGGTGCAAA	GGACCCGTGG	ATGTTTCGGG	TTTCAGCCCC	CCGGGCGCCG
19621	CGCGTTTCCA	GGAAGTACGG	CACCGCCAGC	GCACTACTGC	CGAATATG	CCTACATCCT
19681	TCCATGGCCG	CTACCCCGG	CTATCGTGGC	TACACCTACC	GCCCCAGAAG	ACGAGCGACT
19741	ACCCGACGCC	GAACCACCAC	TGGAACCCGC	CGCGCGGTC	GCCGTGGCCA	GCCCCGTCTG
19801	GCCCCGATTT	CGGTGCGCAG	GGTGGCTGCG	GAAGGAGGCA	GGACCCGTGG	GCTGCCAACCA
19861	GCGCGCTACC	ACCCGAGCAT	CGTTTAAAG	CCGGTCTTTG	TGGTTCCTGC	AGATATGGCC
19921	CTCACCTGCC	GCCTCCGTTT	CCCGGTGCCG	GGATTCCGAG	GAAGAATGCA	CCGTAGAGGG
19981	GGCATGGCCG	GCCACGGCCT	GACGGGCGGC	ATGGGTGGTG	CGCACCCAGG	GCGGGGCGGC
20041	GCGTGGCACC	GTCGCATGCG	CGGCGGTATC	CTGCCCTTCC	TTATTCCACT	GATGGCGCGG
20101	GCGATTGGCG	CGGTGCCCGG	AATTGCATCC	GTGGCCTTGC	AGGCGCAGAG	ACACTGATTA
20161	AAAACAAGTT	GCATGTGGA	AAATCAAAAT	AAAAAGTCTG	GAGTCTCAAG	CTCGCTTGST
20221	CCTGTAACATA	TTTTGTAGAA	TGGAAGACAT	CAACTTTGCG	TCTCTGGCCC	CGCGACACCG
20281	CTCGCGCCCG	TTCATGGGAA	ACTGGCAAGA	TATGGCACC	AGCAATATGA	GCGGTGGCGC
20341	CTTCAGCTGG	GGCTCGCTGT	GGAGCGGCAT	TAAAAATTTT	GGTTCCACCA	TTAAGAACTA
20401	TGGCAGCAAG	GCCTGGAACA	GCAGCACAGG	CCAGATGCTG	AGGGACAAGT	TGAAAGAGCA
20461	AAATTTCCAA	CAAAAGGTGG	TAGATGGCCT	GGCCTCTGGC	ATTAGCGGGG	TGGTGGACCT
20521	GGCCAACCA	GCAGTGCAAA	ATAAGATTAA	CAGTAAGCTT	GATCCCCGCC	CTCCCGTAGA
20581	GGAGCCTCCA	CGGGCGGTGG	AGACAGTGTG	TCCAGAGGGG	CGTGGCGAAA	AGCGTCCCGG
20641	GCCGACAGG	GAAGAACTC	TGCTGAAGCA	AATAGATGAG	CCTCCCTCGT	ACGAGGAGGC
20701	ACTAAAGCAA	GGCCTGCCCA	CCACCCGTCC	CATCGGCCCC	ATGGCTACCG	GAGTGTCTGG
20761	CCAGCACCA	CCTGTAAAGC	TGGACCTGCC	TCCCCCGCT	GACACCCAGC	AGAAACCTGT
20821	GCTGCCAGGG	CCGTCCCGCG	TTGTTGTAAC	CCGCCCTAGC	CGCGGTCCC	TGCGCGGTGC
20881	OGCCAGCGGT	CGCGGATGGA	TGCGGCGCGT	AGCCAGTGGC	AACTGGCAAA	GCACACTGAA
20941	CAGCATGCTG	GGTCTGGGGG	TGCAATCCCT	GAAGCGCGGA	CGATGCTTCT	AAATAGCTAA
21001	CGTGTGCTAT	GTGTCATGTA	TGGTCCATG	TGCGCGCCAG	AGGAGCTGCT	GAGCCCGCGT
21061	CGCGCGCTTT	TCCAAGATGG	CTACCCCTTC	GATGATGCGG	CAGTGGTCTT	ACATGCACAT
21121	CTCGGGCCAG	GACGCTTCGG	AGTACCTGAG	CCCCGGGCTG	GTGCAGTTTG	CCCGCCCCAC
21181	CGAGACGTAC	TTCAGCCTGA	ATAACAAGTT	TAGAAACCCC	ACGGTGGCAC	CTACGCACGA
21241	CGTAACCACA	GACCGGTCCC	AGCGTTTGAC	GCTGCGGTTT	ATCCCTGTGG	ACCGGAGGGA
21301	TACCGGTAC	TGCTACAAAG	CGCGGTTTAC	CCTGGCTGTG	GGTGACAACC	GTGTGCTTGA
21361	TATGGCTTCC	ACGTACTTTG	ACATCCGCGG	CGTCTGGAC	AGGGGGCCTA	CTTTTAAGCC
21421	CTACTCCGGC	ACTGCCTACA	ACGCTCTAGC	TCCCAAGGGC	GCTCCTAACT	CCTGTGAGTG
21481	GGAACAAACC	GAAGATAGCG	GCGGGGAGT	TGCCGAGGAT	GAAGAAGAGG	AAGATGAAGA
21541	TGAAGAAGAG	GAAGAAGAAG	AGCAAAACGC	TGAGATCAG	CCTACTAAGA	AAACACATGT
21601	CTATGCCCCAG	GCTCCTTTGT	CTGGAGAAAC	AAATTACAAA	AGCGGGCTAC	AAATAGGATC
21661	AGACAATGCA	GAAACACAAG	CTAAACCTGT	ATACGCAGAT	CCTTCTTATC	AACCAGAACCC
21721	TCAAATTGGC	GAATCTCAGT	GGAACGAAGC	TGATGCTAAT	CGCGCAGGAG	GGAGAGTGCT
21781	TAAAAAAACA	ACTCCCATGA	AACCATGCTA	TGGATCTTAT	GCCAGGCCTA	CAATCCTTT
21841	TGGTGGTCAA	TCCGTTCTGG	TTCCGGATGA	AAAACGGGTG	CCTCTTCCAA	AGGTTGACTT
21901	GCAATTCTTC	TCAAATACTA	CCTCTTTGAA	CGACCGGCA	GGCAATGCTA	CTAAACCAAA
21961	AGTGGTTTTG	TACAGTGAAG	ATGTAAATAT	GGAACCCCA	GACACACATC	TGCTTTACAA
22021	ACCTGGAATA	GGTGTAGAAA	ATTCTAAAGC	GCGACAATTT	TATGTTGGGT	CAACAATCTA
22081	ACCCAATTAC	ATTGCTTTCA	GCGACAATTT	TATTGGCCTA	ATGTATTATA	ACAGCACTGG
22141	CAACATGGGT	GTTCTTGCTG	GTCAGGCATC	GCAGCTAAAT	GCCGTGGTAG	ATTTGCAAGA
22201	CAGAAACACA	GAGCTGTCTT	ATCAACTCTT	GCTTGATTCC	ATAGGTGATA	GAACCAGATA
22261	TTTTTCTATG	TGGAATCAGG	CTGTAGACAG	CTATGATCCA	GATGTTAGAA	TCATTGAAA
22321	CCATGGAAC	GAGGATGAAT	TGCCAAATTA	TTGTTTTCCT	CTTGGGGGTA	TTGGGGTAAC

## Nucleotide Sequence Analysis (cont.)

22381	TGACACCTAT	CAAGCTATTA	AGGCTAATGG	CAATGGCTCA	GGCGATAATG	GAGATACTAC
22441	ATGGACAAAA	GATGAACTT	TTGCAACACG	TAATCAAATA	GGAGTGGGTA	ACAACCTTGC
22501	CATGGAAATT	AACCTAAATG	CCAACCTATG	GAGAAATTTT	CTTTACTCCA	ATATTGOGCT
22561	GTACCTGCCA	GACAAGCTAA	AATACAACCC	CACCAATGTG	GAAATATCTG	ACAACCCCAA
22621	CACCTACGAC	TACATGAACA	AGCGAGTGGT	GGCTCCCGGG	CTTGTAGACT	GCTACATPAA
22681	CCTTGGGGGG	CGCTGGTCTC	TGCACTACAT	GGACAAOGTT	AATCCCTTTA	ACCACCACCG
22741	CAATCGGGGC	CTCCGTTATC	GCTCCATGTT	GTTGGGAAAC	GGCCGCTACG	TGOCCTTTCA
22801	CATTCAAGTG	CCCCAAAAGT	TTTTTGCCAT	TAAAAACCTC	CTCCTCCTGC	CAGGCTCATA
22861	TACATATGAA	TGGAACCTCA	GGAAGGATGT	TAACATGGTT	CTGCAGAGCT	CTCTGGGAAA
22921	CATCTTTAGA	GTTGACGGGG	CTAGCATTAA	GTTTGACAGC	ATTTGTCTTT	ACGCCACCTT
22981	CTTCCCCATG	GCCCCAACA	CGCCCTCCAC	GCTGGAAGCC	ATGCTCAGAA	ATGACACCAA
23041	CGACCAAGTC	TTTAATGACT	ACCTTTCCGC	CGCCACATG	CTATACCCCA	TACCCGCCAA
23101	CGCCACCAAC	GTGCCCATCT	CCATCCCATC	GCGCAACTGG	GCAGCATTTT	GCGGTTGGGC
23161	CTTCACACGC	TTGAAGACAA	AGGAACCCCC	TTCCCTGGGA	TCAGGCTACG	ACCCTTACTA
23221	CACCTACTCT	GGCTCCATAC	CATACCTTGA	CGGAACCTTC	TATCTTAATC	ACACCTTTAA
23281	GAAGGTGGCC	ATTACCTTTG	ACTCTTCTGT	TAGCTGGCCG	GGCAACGACC	GCCTGCTTAC
23341	TCCCCAATGAG	TTTGAGATTA	AACGCTCAGT	TGACGGGGAG	GGCTACAACG	TAGCTCAGTG
23401	CAACATGACC	AAGGACTGGT	TCCTGGTGCA	GATGTGGGOC	AACATAAATA	TTGGCTACCA
23461	GGGCTTCTAC	ATTCCAGAAA	GCTACAAGGA	CCGCATGTAC	TGGTTCTTCA	GAAACTTCCA
23521	GCCCATGAGC	CGGCAAGTGG	TTGACGATAC	TAAATACAAG	GAGTATCAGC	AGGTTGGAAT
23581	TCTTCACCCAG	CATAACAAC	CAGGATTGCT	AGGCTACCTC	GCTCCCAACA	TGCGCGAGGG
23641	ACAGGCTTAC	CCCGCCAACG	TGCCCTACCC	ACTAATAGGC	AAAACOGCGG	TTGACAGTAT
23701	TACCCAGAAA	AAGTTTCTTT	GCGATCGCAC	CCTTTGGCGC	ATCCCATTTCT	CCAGTAACCTT
23761	TATGTCCATG	GGCGCACTCA	CAGACCTGGG	CCAAAACCTT	CTCTACGCCA	ACTCCGCCCA
23821	CGCGCTAGAC	ATGACTTTTG	AGGTGGATCC	CATGGACGAG	CCCACCCCTT	TTTATGTTTT
23881	GTTTGAAGTC	TTTGACGTGG	TCCGTGTGCA	CCAGCCGCAC	CGCGCGGTCA	TCCAGACCGT
23941	GTACCTGGCG	ACGCCCTTCT	CGCGCGGCAA	CGCCACAACA	TAAAAGAAGC	AAGCAACATC
24001	AACAACAGCT	GCCGCCATGG	GCTCCAGTGA	GCAGGAACCT	AAAGCCATTG	TCAAAGATCT
24061	TGGTGTGGG	CCATATTTTT	TGGGCACCTA	TGACAAAGCG	TTTCCAGGCT	TTGTTTCTCC
24121	ACACAAGCTC	GCCTGCGCCA	TAGTCAATAT	GGCCGGTGGC	GAGACTGGGG	GCGTACACTG
24181	GATGGCCTTT	GCCTGGAACC	CGCGCTCAAA	AACATGCTAC	CTCTTTGAGC	CCTTTGCTTT
24241	TTCTGAACCA	CGACTCAAGC	AGGTTTACCA	GTTTGAGTAC	GAGTCACTCC	TGCGCCGTAG
24301	CGCCATTGCT	TCTTCCCCCG	ACCGCTGTAT	AACGCTGGAA	AAGTCCACCC	AAAGCGTGCA
24361	GGGGCCCAAC	TGGGCGCGCT	GTGGACTATT	CTGCTGCATG	TTTCTCCACG	CCTTTGCCAA
24421	CTGCCCCCAA	ACTCCCATGG	ATCACAACCC	CACCATGAAC	CTTATTACCG	GGGTACCCAA
24481	CTCCATGCTT	AACAGTCCCC	AGGTACAGCC	CACCCCTGGT	CCCAACCAGG	AACAGCTCTA
24541	CAGCTTCTCT	GAGCGCCACT	CGCCCTACTT	CCGCAGCCAC	AGTGGCGAGA	TTAGGACCGC
24601	CACCTCTTTT	TGTCACCTGA	AAAACATGTA	AAAATATGTT	ACTAGGAGAC	ACTTTCAATA
24661	AAAGCAAAATG	TTTTTATTTG	TACACTCTCG	GGTGATTATT	TACCCCCAC	CCTTGCCTTC
24721	TGCGCGGTTT	AAAAATCAAA	GGGGTTCTGC	CGCGCATGCG	TATGCGCCAC	TGGCAGGGAC
24781	ACGTGCGGAT	ACTGGTGTTT	AGTGCTCCAC	TAAACTCAG	GCACAACCAT	CCGCGGCAGC
24841	TGGTGAAGT	TTTCACTCCA	CAGGCTGGCG	ACCATCACCA	ACCGGTTTAG	CAGGTGGGGC
24901	GCCGATATCT	TGAAGTCGCA	GTTGGGGCCT	CCGCCCTGGG	CGCGGAGTTT	GCGATACACA
24961	GGGTGCGAGC	ACTGGAACAC	TATCAGCGCC	GGGTGGTGCA	CGCTGGCCAG	CACGCTCTTG
25021	TGGGAGATCA	GATCCGCGTC	CAGGTCCCTCC	CGGTTGCTCA	GGGGAACCGG	AGTCAACTTT
25081	GGTAGCTTTC	TTCCCAAAAA	GGGTGCATGC	CCAGGCTTTG	AGTTGCACTC	GCACCGTAGT
25141	GGCATCAGAA	GGTGACCGTG	CCCGGTCTGG	CGGTTAGGAT	ACAGCGCCTG	CATGAAAGCC
25201	TTGATCTGCT	TAAAAGCCAC	CTGAGCCTTT	GCGCCTTACG	AGAAGAACAT	GCCGCAAGAC
25261	TTGCCGGAAT	ACTGATTGGC	CGGACAGGCC	GCGTCATGCA	CGCAGCACCT	TGCGTGGGTG
25321	TTGGAGATCT	GCACCACATT	TGGGCCCCAC	CGGTTCTTCA	CGATCTTGGC	CTTGCTAGAC
25381	TGCTCCTTCA	GCGCGCGCTG	CCCGTTTTCG	CTCGTCAAT	CCATTTCAAT	CACGCTGCTC
25441	TTATTTATCA	TAATGCTCCC	GTGTAGACAC	TTAAGCTCGC	CTTCGATCTC	AGCGCAGCGG
25501	TGCAGCCACA	ACGCGCAGCC	CGTGGGCTCG	TGGTGTCTGT	AGGTTACCTC	TGCAACGAC
25561	TGCAGGTACG	CCTGCAGGAA	TCCGCCCATC	ATCGTCACAA	AGGCTCTGTT	GCTGGTGAAG
25621	GTCAGCTGCA	ACCCGCGGTG	CTCCTCGTTT	AGCCAGGTCT	TGCATACGGC	CGCCAGAGCT
25681	TCCACTTGGT	CAGGCAGTAG	CTTGAAGTTT	GCCTTTAGAT	CGTTATCCAC	GTGGTACTTG
25741	TCCATCAACG	CGCGCGCAGC	CTCCATGCCC	TTCTCCACAG	CAGACACGAT	CGGCAGGCTC

## Nucleotide Sequence Analysis (cont.)

25801 AGCGGGTTTA TCACCGTGCT TTCACTTTCC GCTTCACTGG ACTCTTCCTT TTCCCTCTTGC  
 25861 GTCGCGATAC CCGCGGCCAC TGGGTGCTCT TCATTTCAGCC GCCGCACCGT GCGCTTACCT  
 25921 CCCTTGCCGT GCTTGATTAG CACCGGTGGG TTGCTGAAAC CCACCATTTG TAGCGCCACA  
 25981 TCTTCTCTTT CTTCCTCGCT GTCCACGATC ACCTCTGGGG ATGGGGGGCG CTCGGGCTTG  
 26041 GGAGAGGGGC GCTTCTTTTT CTTTTGGGAC GCAATGGCCA AATCGCCCGT CGAGGTCGAT  
 26101 GGCGCGGGGC TGGGTGTGCG CGGCACCAGC GCATCTTGTG ACGAGTCTTC TTGCTCTCG  
 26161 GACTCCAGAC GCCGCTCAG CGCTTTTTT GGGGGCGCG GGGGAGCGG CGCGAAGGC  
 26221 GACGGGAGC ACACGTCCTC CATGCTTGGT GGAAGTGGG CGGCACCGCG TCCGCGCTCG  
 26281 GGGGTGGTTT CGCGCTGCTC CTCTTCCCGA CTGCCCATT CTCTCTCTA TAGGCAGAAA  
 26341 AAGATCATGG AGTCAGTCGA GAAGGAGGAC AGCCTAACCG CCCCCTTTGA GTTGGCCACC  
 26401 ACGGCTTCCA CCGATGCCGC CAACGCGCCT ACCACCTTCC CCGTCGAGGC ACCCGCGCTT  
 26461 GAGGAGGAGG AAGTGATTAT CGAGCAGGAC CCAGGTTTTG TAAGCGAAGA CGACGAGGAT  
 26521 CGCTCAGTAC CAACAGAGGA TAAAAAGCAA GACCAGGAGC ACGCAGAGGC AAACGAGGAA  
 26581 CAAGTCGGGC GGGGGGACCA AAGGCATGCG GACTACCTAG ATGTGGGAGA CGAGCTGCTG  
 26641 TTGAAGCATC TGCAGCGCCA GTGCGCCATT ATCTGCGAGC CGTTGCAAGA GCGCGAAGGC  
 26701 GTGCCCCCTG CCATAGCGGA TGTACGCTT GCCTACGAAC GCCACCTGTT CTCAACCGGC  
 26761 GTACCCCCCA AAGCGCAAGA AAACGCGACA TGCAGGCCCA ACCCGCGCCT CAACTTCTAC  
 26821 CCGTATTATG CCGTGCCAGA GGTGCTTGCC ACCTATCACA TCTTTTCCA AAAGTGCAG  
 26881 ATACCCCTAT CCTGCGGTGC CAACCGCAGC CGAGCGGACA AGCAGCTGGC CTGCGGCGAG  
 26941 GCGCTGTCA TACCTGATAT CGCTCGCTC GACGAAGTGC CAAAAATCTT TGAGGGTCTT  
 27001 GGAAGCGAGC AGAAAAGCGC GGCAAACGCT CTGCAACAAG AAAACAGCGA AAATGAAAGT  
 27061 CACTGTGGAG TGCTGGTGGG ACTTGAGGGT GACAAACGCG CCCTAGCCGT GCTGAAACGC  
 27121 AGCATCGAGG TCACCCACTT TGCTTACCG GCACCTAAC TACCCCCCAA GGTATGAGC  
 27181 ACAGTCATGA GCGAGCTGAT CGTGGCGGCT GCACGACCCC TGGAGAGGGA TGCAAACTTG  
 27241 CAAGAACAAA CCGAGGAGGG CCTACCCGCA GTTGGCGATG AGCAGCTGGC GCGCTGGCTT  
 27301 GAGAAGCGCG AGCCTGCGGA CTGAGGAGG CGAAGCAAGC TAATGATGCG CGCAGTCTT  
 27361 GTTACCGTGG AGCTTGAGTG CATGAGCGG TTCTTTGCTG ACCCGGAGAT GCAGCGCAAG  
 27421 CTAGAGGAAA CGTTGCACTA CACCTTTCCG CAGGCTAAG TGGCGCAGGC CTGCAAAAT  
 27481 TCCAACGTTG AGCTCTGCAA CCTGCTCTCC TACCTTGGA TTTTGACGA AAACCGCTC  
 27541 GGGCAAAACG TGCTTCATTC CACGCTCAAG GCGAGGGCG GCGCGACTA CGTCCGCGAC  
 27601 TGCGTTTACT TATTTCTGTG CTACACCTGG CAAACGGCCA TGGGCGTGTG GCAGCAATGC  
 27661 CTGGAGGAGC GCAACCTAAA GGAGCTGCGA AAGCTGCTAA AGCAAACTT GAAGGACCTA  
 27721 TGGACGGCCT TCAACGAGCG CTCGTGTGCC CGGCACCTGG CGGACATTAT CTTCCCGAA  
 27781 CGCTGCTTA AAACCTGCA ACAGGGTCTG CCAGACTTCA CCAGTCAAAG CATGTTGCAA  
 27841 AACTTTAGGA ACTTTATCCT AGAGGCTTCA GGAATTTCTG CCGCCACCTG CTGTGCGCTT  
 27901 CCTAGCGACT TTGTGCCCCAT TAAGTACCGT GAATGCCCTC CGCGCTTTG GGGTCACTGC  
 27961 TACCTTCTGC AGCTAGCCAA CTACCTTGCC TACCCTCCG ACATCATGGA AGACGTGAGC  
 28021 GGTGACGGCC TACTGGAGTG TCACTGTGCG TGCAACCTAT GCACCCGCA CCGCTCCCTG  
 28081 GTCTGCANTT CGCAACTGCT TAGCGAAAGT CAAATTATCG GTACCTTTGA GCTGCAGGCT  
 28141 CCCTCGCCTG ACGAAAAGTC CGCGGCTCCG GGGTTGAAAC TCACTCCGGG GCTGTGGAGC  
 28201 TCGGCTTACC TTGCAAAATT TGTACCTGAG GACTACCAG CCCACGAGAT TAGGTTCTAC  
 28261 GAAGACCAAT CCGCCCGGCC AAATGCGGAG CTTACCGCCT GCGTCATTAC CCAGGGCCAC  
 28321 ATCCTTGGCC AATTGCAAGC CATCAACAAA GCGCGCCAAG AGTTTCTGCT ACGAAAGGGA  
 28381 CGGGGGGTTT ACCTGGACCC CCAGTCCGCG GAGGAGCTCA ACCCAATCCC CCGCGCGCG  
 28441 GAGCCCTATC AGCAGCCGCG GCGCCTTGCT TCCCAGGATG GCACCCAAAA AGAAGCTGCA  
 28501 CGTGCGCGCG CCGTACCCCA CGGACGAGGA GGAATACTGG GACAGTCAGG CAGAGGAGGT  
 28561 TTTGACGAG GAGGAGGAGA TGTGGAAGA CTGGGACAGC CTAGACGAAG CTTCGAGGC  
 28621 CGAAGAGGTG TCAGACGAAA CACCGTCACC CTGGGTGCGA TTCCCTCCG CGGCGCCCCA  
 28681 GAAATTGGCA ACGGTTCCCA GCATGCTTAC AACCTCCGCT CCTCAGGCGC CGCCGGCACT  
 28741 GCCTGTTTCC CGACCCAAAC GTAGATGGGA CACCACTGGA ACCAGGGCGG GTAAGTCTAA  
 28801 CAGCGCGCG CGGTTAGCCC AAGAGCAACA ACAGCGCAA GGTACCGCT CCGTGGCGCG  
 28861 GCACAAGAAC GCCATAGTTG CTTGCTTGCA AGACTGTGGG GCGAACATCT CTTGCGCCCG  
 28921 CGCTTTCTT CTCTACCATC ACGGCGTGGC CTTCCTCCG AACATCTGCG ATTACTACCG  
 28981 TCATCTCTAC AGCCCCTACT GCACCGCGCG CAGCGGCAGC GGCAGCAACA GCAGCGGTCA  
 29041 CACAGAAGCA AAGGCGAGCG GATAGCAAGA CTGTGACAAA GCCCAAGAAA TCCACAGCGG  
 29101 CGGCAGCAGC AGGAGGAGGA GCGCTGCGTC TGGCGCCCAA CGAACCCGTA TCGACCCGCG  
 29161 AGCTTAGAAA TAGGATTTTT CCCACTCTGT ATGCTATATT TCAACAAAGC AGGGGCCAAG

## Nucleotide Sequence Analysis (cont.)

```

29221 AACAAAGAGCT GAAAATAAAA AACAGGTCTC TGCGCTCCCT CACCGGCAGC TGCCTGTATC
29281 ACAAAGCGA AGATCAGCTT CGGCGCAOGC TGGAGAGCGC GGAGGCTCTC TTCAGCAAAT
29341 ACTGCGCGCT GACTCTTAAG GACTAGTTTC GCGCCCTTTC TCAAATTTAA GCGCGAAAAC
29401 TAGTCATCT CCAGCGGCCA CACCGGGGCG CAGCACCTGT CGTCAGCGCC ATTATGAGCA
29461 AGGAAATTCC CACGCCCTAC ATGTGGAGTT ACCAGCCACA AATGGGACTT GGGGCTGGAG
29521 CTGCCCAAGA CTACTCAACC CGAATAAACT ACATGAGGCG GGAACCCAC ATGATATCCC
29581 GGGTCAAACG AATCCGGGCC CACCGAAACC GGGACATTTC GAATTCCTCT CGAACAGGCG GCTATTACCA
29641 CCACACCTCG TAATAACCTT AATCCCCGTA GTTGGCCGCG TGCCCTGGTG TACCAGGAAA
29701 GTCCCGCTCC CACCACTGTG GTACTTCCCA GAGAGGCCCA GGCGAAGTT CAGATGACTA
29761 ACTCAGGGGC GCAGCTTGCG GCGGGCTTTC GTACAGGGT GGGTGGCCC GGGCAGGGTA
29821 TAACACCTT GAAAATCAGA GGGCGAGGTA TTCAGCTCAA CGACGAGTCG GTGAGCTCCT
29881 CTCCTGGTCT CGTCCGGAC GGGACATTTC AGATCGGCGG CGCTGGCCGC TCTTCATTTA
29941 CGCCCCGTCA GGGATCCTA ACTCTGCAGA CCTCGTCTCT GGAGCGGCGC TCCGAGGCCA
30001 TTGGAACCTT ACAATTTATT GAGGAGTTTG TGCTTCCGGT TTAATTCATC CCCTTTCTGT
30061 GACCTCCCGG CCACTACCCG GACCAGTTTA TTCCCAACTT TGAAGCGGTG AAAGACTCGG
30121 CGGACGGCTA CGACTGAATG ACCAGTGGAG AGGCAGAGCG ACTGCGCCTG ACACACCTCG
30181 ACCACTGCGG CGGCCACAAG TGCTTTGGCC GCGGCTCCGG TGAGTTTGTG TACTTTGAAT
30241 TGCCCGAAGA GCATATCGAG GGCCCGGCGC ACGGCGTCCG GCTCACCACC CAGGTAGAGC
30301 TTACACGTAG CCTGATTGGG GAGTTTACCA AGCGCCCCCT GCTAGTGGAG CGGGAGCGGG
30361 GTCCCTGTGT TCTGACCGTG GTTTGCAACT GTCTTAACCC TGGATTACAT CAAGATCTTT
30421 CTGTGCATCT CTGTGCTGAG TATAATAAAT ACAGAAATTA GAATCTACTG GGGCTCTGTG
30481 CGCCTTCCTG TGAACGCCAC CGTTTTTACC CACCCAAAGC AGACCAAAGC AAACCTCACC
30541 TCCGCTTTGC ACAAGCGGGC CAATAAGTAC CTTACCTGGT ACTTTAAGCG CTCTTCATTT
30601 GTAATTTACA ACAGTTTCCA GCGAGACGAA GTAAGTTTGC CACACAACCT TCTCGGCTTC
30661 AACTACACCG TCAAGAAAAA CACCACCACC ACCACCTCCG TCACCTGCGG GGAACGTACG
30721 AGTGCGTCAC CGGTTGCTGC GCCACACCT ACAGCCTGAG CGTAACCAGA CATTACTCCC
30781 ATTTTTCCAA AACAGGAGGT GAGCTCAACT CCGGAACCTC AGGTCAAAAA AGCATTTTGC
30841 GGGGTGCTGG GATTTTTTAA TTAAGTATAT GAGCAATTCA AGTAACCTTA CAAGCTTGTC
30901 TAATTTTTCT GGAATTGGGG TCGGGGTAT CCTACTCTT GTAATTCGTG TTATCTTAT
30961 ACTAGCACTT CTGTGCCCTA GGGTTGCCGC CTGCTGCACG CACGTTTGTA CCTATGTGTA
31021 GCTTTTTTAA CGCTGGGGGC AACATCCAAG ATGAGGTACA TGATTTTAGG CTTGCTCGCC
31081 CTTCGCGCAG TCTGCAGGCG TGCCAAAAAG GTTGAGTTTA AGGAACCAGC TTGCAATGTT
31141 ACATTTAAAT CAGAAGCTAA TGAATGCACT ACTCTTATAA AATGCACCAC AGAACATGAA
31201 AAGCTTATTA TTGCCACAA AGACAAAATT GGCAAGTAG CTGTATATGC TATTTGGCAG
31261 CCAGGTGACA CTAACGACTA TAATGTCACA GTCTTCCAAG GTGAAAAATG TAAACTTTT
31321 ATGTATAAAT TTCCATTTTA TGAAATGTGC GATATTACCA TGTACATGAG CAAACAGTAC
31381 AAGTTGTGGC CCCCAAAAA GTGTTTAGAG AACACTGGCA CCTTTTGTTC CACCGCTCTG
31441 CTTATTACAG CGCTTGCTTT GGTATGTACC TTACTTTATC TCAAATACAA AAGCAGACCG
31501 AGTTTATTG ATGAAAAGAA AATGCCCTGA TTTCCGCTT GCTTGTATTC CCCTGGACAA
31561 TTTACTCTAT GTGGGATATG CTCCAGGCGG GCAAGATTAT ACCCACAACC TTCAAATCAA
31621 ACTTTCCTGG ACCTTAGGCG CTGATTTCTG CCAGCGCCTG CACTGCAAAT TTGATCAAA
31681 CCAGCTTCAG CTTGCCTGCT CCAGAGATGA CCGGCTCAAC CATCGCGCCC ACAACGGACT
31741 ATCGCAACAC CACTGCTACC GGACTAACAT CTGCCCTAAA TTTACCCCAA GTTCATGCCT
31801 TTGTCAATGA CTGGGCGAGC TTGGACATGT GGTGGTTTTC CATAGCGCTT ATGTTTGTGT
31861 GCCTTATTAT TATGTGGCTT ATTTGTTGCC TAAAGCGCAG ACGCGCCAGA CCCCCCATCT
31921 ATAGGCTTAT CATGTGCTC AACCCACACA ATGAAAAAAT TCATAGATTG GACGCTCTGA
31981 AACCATGTTT TCTTCTTTTA CAGTATGATT AAATGAGACA TGATTCTCG AGTTCTTATA
32041 TTATTGACCC TTGTTGCGCT TTTCTGTGCG TGCTCTACAT TGGCCGCGGT CGCTCACATC
32101 GAAGTAGATT GCATCCCAAC TTTCACAGTT TACCTGCTTT ACGGATTGTG CACCTTATC
32161 CTCATCTGCA GCCTCGTCACT TGTAGTCATC GCCTTCATTC AGTTCAATGA CTGGGTTTGT
32221 GTGGCATTG CGTACCTCAG GCACCATCG CAATACAGAG ACAGGACTAT AGCTGATCTT
32281 CTCAGAAATC TTTAATTATG AAAACGGAGT TCATTTTGTG TTTGCTGATT TTTTGGCCCC
32341 TACCTGTGCT TTGCTCCCAA ACCTCAGGCG CTCCAAAAAG ACATATTTCC TGCAATTTCA
32401 CTCAAATATG GAACATTTCC AGCTGCTACA ACAAACAGAG CGATTTGTCA GAAGCCTGCT
32461 TATACGCCAT CATCTCTGTC ATGGTTTITT GCAGTACCAT TTTTGGCCCTA GCCATATATC
32521 CATACCTTGA CATGGCTGG AATGCCATAG ATGCCATGAA CCACCTACT TTCCAGTGC
32581 CCGCTGTCAT ACCACTGCAA CAGGTTATTG CCCCATCAA TCAGCCTCGC CCCCCTTCTC

```



## Nucleotide Sequence Analysis (cont.)

```

32641 CCACCCCCAC TGAGATTAGC TACTTTAATT TGACAGGTGG AGATGACTGA ATCTCTAGAT
32701 CTAGAATTGG ATGGAATTAA CACCGAACAG CGCCTACTAG AAAGGCGCAA GCGCGCGTCC
32761 GAGCGAGAAC GCCTAAACAA AGAAGTTGAA GACATGGTTA ACCTACACCA GTGTAAAGA
32821 GGTATCTTTT GTGTGGTCAA GCAGGCCAAA CTACCTAAG AAAAAACCAC TACCGGCAAC
32881 CGCCTCAGCT ACAAGCTACC CACCCAGGCG CAAAACTGG TCCTTATGGT GGGAGAAAAA
32941 COTATCACCG TCACCCAGCA CTCGGCAGAA ACAGAGGGCT GCCTGCACCT CCCCTATCAG
33001 GGTCCAGAGG ACCTCTGCAC TCTTATTAAA ACCATGTGTG GTATTAGAGA TCTTATTCCA
33061 TTCAACTAAC ATAAACACAC AATAAATTAC TTAATAAAA TCAGTCAGCA AATCTTTGTC
33121 CAGCTTATTC AGCATCACCT CCTTCTCTTC CTCCCAACTC TGGTATCTCA GCCGCCTTTT
33181 AGCTGCAAAC TTTCTCCAAA GTTTAAATGG GATGTCAAAT TCCTCATGTT CTGTCCCTC
33241 CGCACCACCT ATCTTCATAT TGTTCAGAT GAAACGGGCC AGACCGTCTG AAGACACCTT
33301 CAACCCCGTG TATCCATATG ACACAGAAAC CGGCCCTCCA ACTGTGCCCT TCTTACCCC
33361 TCCATTTGTT TCACCAATG GTTTCCAGA AAGTCCCCCT GGAGTTCTCT CTCTAOGGT
33421 CTCCGAACCT TTGGACACCT CCCACGGCAT GCTTGOGCTT AAAATGGGCA GGGGTCTTAC
33481 CCTAGACAAAG GCCGGAACCC TCACCTCCCA AAATGTAACC ACTGTFACTC AGCCACTTAA
33541 AAAAAAQAAG TCAAACATAA GTTTGGACAC CTCGCGACCA CTTACAATTA CCTCAGGCGC
33601 CCTAACAGTG GCAACCAACCG CTCCTCTGAT AGTTACTAGC GCGCTCTTA GCGTACAGTC
33661 ACAAGCCCCA CTGACCGTGC AAGACTCCAA ACTAAGCATT GCTACTAAG GGGCCATTAC
33721 AGTGTTCAGAT GGAAAGCTAG CCCTGCAAAC ATCAGCCCCC CTCTCTGGCA GTGACAGCGA
33781 CACCCTTACT GTAAGTGCAT CACCCCGCT AACTACTGCC ACGGGTAGCT TGGGCATTAA
33841 CATGGAAGAT CCTATTATG TAAATAATGG AAAAATAGGA ATTAATAA GCGGTCCTTT
33901 GCAAGTAGCA CAAAACCTCG ATACACTAAC AGTAGTTACT GGACCAGGTG TCACCGTTGA
33961 ACAAACTCC CTTAGAACCA AAGTTGCAGG AGCTATTGGT TATGATTAT CAAACAACAT
34021 GGAAATTTAA ACGGGCGGTG GCATGCGTAT AAATAACAAC TTGTTAATTC TAGATGTGGA
34081 TTACCCATTT GATGCTCAA CAAACTACG TCTTAACTG GGGCAGGAC CCCTGTATAT
34141 TAATGCATCT CATAACTTGG ACATAAACTA TAACAGAGGC CTATACCTTT TTAATGCATC
34201 AAACAATACT AAAAACTGG AAGTTAGCAT AAAAAATCC AGTGGACTAA ACTTTGATAA
34261 TACTGCCATA GCTATAAATG CAGGAAAGGG TCTGGAGTTT GATACAAACA CATCTGAGTC
34321 TCCAGATATC AACCCTAATA AACTTAAAT TGGCTCTGGC ATTGATTACA ATGAAAACGG
34381 TGCCATGATT ACTAACTTG GAGCGGCTT AAGCTTTGAC AACTCAGGGG CCATTACAAT
34441 TGGAACAAA AATGATGACA AACTTACCT GTGGACAACC CCAGACCCAT CTCCTAAGTC
34501 CAGAAATCAT TCAGATAATG ACTGCAAAAT TACTTTGGTT CTTACAAAT GTGGGAGTCA
34561 AGTACTAGCT ACTGTAGCTG CTTTGGCTGT ATCTGCAGAT CTTTCATCCA TGACAGGCAC
34621 CGTTGCAAGT GTTAGTATAT TCCTTAGATT TGACAAAAC GGTGTTCTAA TGGAGAATC
34681 CTCACCTAAA AAACATTACT GGAACCTTAG AAATGGGAAC TCAACTAATG CAAATCCATA
34741 CACAAATGCA GTTGGATTTA TGCCTAACCT TCTAGCCTAT CCAAAAACCC AAAGTCAAAC
34801 TGCTAAAAAT AACATTGTCA GTCAAGTTTA CTTCATGGT GATAAACTA AACCTATGAT
34861 ACTTACCATT ACACCTAATG GCACTAGTGA ATCCACAGAA ACTAGCGAGG TAAGCACTTA
34921 CTCTATGTCT TTTACATGGT CCTGGGAAAG TGGAAAATAC ACCACTGAAA CTTTGTCTAC
34981 CAACTCTTAC ACCTTCTCCT ACATTGCCCA GGAATAAAGA ATCGTGAACC TGTTCATGTT
35041 TATGTTTCAA CGTGGGATCC TTTATTATAG GGGAAAGTCCA CGCCTACATG GGGGTAGAGT
35101 CATAATCGTG CATCAGGATA GGGCGGTGGT GCTGCAGCAG CGCGCAATA AACTGCTGCC
35161 GCCGCGCTC CGTCTGCAG GAATACAACA TGGCAGTGGT CTCCTCAGCG ATGATTGCGA
35221 CCGCCCGCAG CATGAGACGC CTTGTCTCC CTGGCACAGCA GCGCACCTG ATCTCACTTA
35281 AATCAGCACA GTAAGTGCAG CACAGCACCA CAATATTGTT CAAAATCCCA CAGTGCAAGG
35341 CGCTGTATCC AAAGCTCATG GCGGGGACCA CAGAACCCAC GTGGCCATCA TACCACAAGC
35401 GCAGGTAGAT TAAGTGGGGA CCCCTCATAA ACAGCTGGA CATAAACATT ACCCTCTTTG
35461 GCATGTTGTA ATTCAACACC TCCCGGTACC ATATAAACCT CTGATTAAAC ATGCGGCCAT
35521 CCACCACCAT CCTAAACCAG CTGGCCAAAA CCTGCCCGCC GGCTATGCAC TGCAGGGAAC
35581 CGGACTGGA ACAATGACAG TGGAGAGCCC AGGACTCGTA ACCATGGATC ATCATGCTCG
35641 TCAATGATATC AATGTTGGCA CAACACAGGC ACAGTGCAT AACTTCCTC AGGATTACAA
35701 GCTCCTCCCG CGTCAGAACC ATATCCACG GAACAACCCA TTCCTGAATC AGCGTAAATC
35761 CCACACTGCA GGAAGACCT CGCACGTAC TCAGTTGTG CATTGTCAA GTGTACATT
35821 CGGCGAGCAG CGGATGATCC TCCAGTATGG TAGCGCGGGT CTCGTCTCTA AAAGGAGGTA
35881 GCGGATCCCT ACTGTACGGA GTGCGCCGAG ACAACCGAGA TCGTGTGGT CGTAGTGTCA
35941 TGCCAAATGG AACGCCGAG GTAGTCATAT TTCATCGACA CCGCACAGC TCAATCAGTC
36001 ACAGTGTAAG AAGGGCCAAG TACAGAGCCA GTATATATAG GACTAAAAAA TGACGTAACG

```

**Nucleotide Sequence Analysis (cont.)**

36061 GTTAAAGTCC ACAAAAACA CCCAGAAAC CGCACGOGAA CCTACGCCCA GAAACGAAAG  
36121 CCAAAAACC CACAACTTCC TCAAATCTTC ACTTCGGTTT TCCCACGATA CGTCACTTCC  
36181 CATTITAAAA AACTACAAT TOCCAATACA TGCAAGTTAC TCGCCCTAA AACCTACGTC  
36241 ACCCGCCCCG TTCCCACGCC CCGCGCCACG TCACAAACTC CACCCCTCA TTATCATATT  
36301 GGCTTCAATC CAAAATAAGG TATATTATGA TGATG

//

- 96 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- 5
- (i) APPLICANTS: Gregory, R.J., Armentano, D., Couture, L.A., Smith, A.E.
- 10
- (ii) TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
- (iii) NUMBER OF SEQUENCES: 9
- 15
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: LAHIVE & COCKFIELD
- (B) STREET: 60 STATE STREET, SUITE 510
- (C) CITY: BOSTON
- (D) STATE: MASSACHUSETTS
- (E) COUNTRY: USA
- 20
- (F) ZIP: 02109
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk
- 25
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: ASCII
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 30
- (B) FILING DATE: 02-DEC-1993
- (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: US 07/985,478
- 35
- (B) FILING DATE: 02-DEC-1992
- (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Hanley, Elizabeth A.
- 40
- (B) REGISTRATION NUMBER: 33,505
- (C) REFERENCE/DOCKET NUMBER: NZI-014CP2PC
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: (617) 227-7400
- 45
- (B) TELEFAX: (617) 227-5941

## (2) INFORMATION FOR SEQ ID NO:1:

- 50
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6129 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- 55
- (ii) MOLECULE TYPE: cDNA

- 97 -

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 133..4572

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

10	AATTGGAAGC AAATGACATC ACAGCAGGTC AGAGAAAAAG GGTGAGCGG CAGGCACCCA	60
	GAGTAGTAGG TCTTTGGCAT TAGGAGCTTG AGCCCAGACG GCCCTAGCAG GGACCCCAGC	120
	GCCCGAGAGA CC ATG CAG AGG TCG CCT CTG GAA AAG GCC AGC GTT GTC	168
15	Met Gln Arg Ser Pro Leu Glu Lys Ala Ser Val Val	
	1 5 10	
	TCC AAA CTT TTT TTC AGC TGG ACC AGA CCA ATT TTG AGG AAA GGA TAC	216
	Ser Lys Leu Phe Phe Ser Trp Thr Arg Pro Ile Leu Arg Lys Gly Tyr	
20	15 20 25	
	AGA CAG CGC CTG GAA TTG TCA GAC ATA TAC CAA ATC CCT TCT GTT GAT	264
	Arg Gln Arg Leu Glu Leu Ser Asp Ile Tyr Gln Ile Pro Ser Val Asp	
	30 35 40	
25	TCT GCT GAC AAT CTA TCT GAA AAA TTG GAA AGA GAA TGG GAT AGA GAG	312
	Ser Ala Asp Asn Leu Ser Glu Lys Leu Glu Arg Glu Trp Asp Arg Glu	
	45 50 55 60	
30	CTG GCT TCA AAG AAA AAT CCT AAA CTC ATT AAT GCC CTT CGG CGA TGT	360
	Leu Ala Ser Lys Lys Asn Pro Lys Leu Ile Asn Ala Leu Arg Arg Cys	
	65 70 75	
35	TTT TTC TGG AGA TTT ATG TTC TAT GGA ATC TTT TTA TAT TTA GGG GAA	408
	Phe Phe Trp Arg Phe Met Phe Tyr Gly Ile Phe Leu Tyr Leu Gly Glu	
	80 85 90	
40	GTC ACC AAA GCA GTA CAG CCT CTC TTA CTG GGA AGA ATC ATA GCT TCC	456
	Val Thr Lys Ala Val Gln Pro Leu Leu Gly Arg Ile Ile Ala Ser	
	95 100 105	
	TAT GAC CCG GAT AAC AAG GAG GAA CGC TCT ATC GCG ATT TAT CTA GGC	504
	Tyr Asp Pro Asp Asn Lys Glu Glu Arg Ser Ile Ala Ile Tyr Leu Gly	
	110 115 120	
45	ATA GGC TTA TGC CTT CTC TTT ATT GTG AGG ACA CTG CTC CTA CAC CCA	552
	Ile Gly Leu Cys Leu Leu Phe Ile Val Arg Thr Leu Leu Leu His Pro	
	125 130 135 140	
50	GCC ATT TTT GGC CTT CAT CAC ATT GGA ATG CAG ATG AGA ATA GCT ATG	600
	Ala Ile Phe Gly Leu His His Ile Gly Met Gln Met Arg Ile Ala Met	
	145 150 155	
55	TTT AGT TTG ATT TAT AAG AAG ACT TTA AAG CTG TCA AGC CGT GTT CTA	648
	Phe Ser Leu Ile Tyr Lys Lys Thr Leu Lys Leu Ser Ser Arg Val Leu	
	160 165 170	

	GAT AAA ATA AGT ATT GGA CAA CTT GTT AGT CTC CTT TCC AAC AAC CTG	696
	Asp Lys Ile Ser Ile Gly Gln Leu Val Ser Leu Leu Ser Asn Asn Leu	
	175 180 185	
5	AAC AAA TTT GAT GAA GGA CTT GCA TTG GCA CAT TTC GTG TGG ATC GCT	744
	Asn Lys Phe Asp Glu Gly Leu Ala Leu Ala His Phe Val Trp Ile Ala	
	190 195 200	
10	CCT TTG CAA GTG GCA CTC CTC ATG GGG CTA ATC TGG GAG TTG TTA CAG	792
	Pro Leu Gln Val Ala Leu Leu Met Gly Leu Ile Trp Glu Leu Leu Gln	
	205 210 215 220	
15	GCG TCT GCC TTC TGT GGA CTT GGT TTC CTG ATA GTC CTT GCC CTT TTT	840
	Ala Ser Ala Phe Cys Gly Leu Gly Phe Leu Ile Val Leu Ala Leu Phe	
	225 230 235	
20	CAG GCT GGG CTA GGG AGA ATG ATG ATG AAG TAC AGA GAT CAG AGA GCT	888
	Gln Ala Gly Leu Gly Arg Met Met Met Lys Tyr Arg Asp Gln Arg Ala	
	240 245 250	
25	GGG AAG ATC AGT GAA AGA CTT GTG ATT ACC TCA GAA ATG ATT GAA AAT	936
	Gly Lys Ile Ser Glu Arg Leu Val Ile Thr Ser Glu Met Ile Glu Asn	
	255 260 265	
30	ATC CAA TCT GTT AAG GCA TAC TGC TGG GAA GAA GCA ATG GAA AAA ATG	984
	Ile Gln Ser Val Lys Ala Tyr Cys Trp Glu Glu Ala Met Glu Lys Met	
	270 275 280	
35	ATT GAA AAC TTA AGA CAA ACA GAA CTG AAA CTG ACT CGG AAG GCA GCC	1032
	Ile Glu Asn Leu Arg Gln Thr Glu Leu Lys Leu Thr Arg Lys Ala Ala	
	285 290 295 300	
40	TAT GTG AGA TAC TTC AAT AGC TCA GCC TTC TTC TTC TCA GGG TTC TTT	1080
	Tyr Val Arg Tyr Phe Asn Ser Ser Ala Phe Phe Phe Ser Gly Phe Phe	
	305 310 315	
45	GTG GTG TTT TTA TCT GTG CTT CCC TAT GCA CTA ATC AAA GGA ATC ATC	1128
	Val Val Phe Leu Ser Val Leu Pro Tyr Ala Leu Ile Lys Gly Ile Ile	
	320 325 330	
50	CTC CGG AAA ATA TTC ACC ACC ATC TCA TTC TGC ATT GTT CTG CGC ATG	1176
	Leu Arg Lys Ile Phe Thr Thr Ile Ser Phe Cys Ile Val Leu Arg Met	
	335 340 345	
55	GCG GTC ACT CGG CAA TTT CCC TGG GCT GTA CAA ACA TGG TAT GAC TCT	1224
	Ala Val Thr Arg Gln Phe Pro Trp Ala Val Gln Thr Trp Tyr Asp Ser	
	350 355 360	
60	CTT GGA GCA ATA AAC AAA ATA CAG GAT TTC TTA CAA AAG CAA GAA TAT	1272
	Leu Gly Ala Ile Asn Lys Ile Gln Asp Phe Leu Gln Lys Gln Glu Tyr	
	365 370 375 380	
65	AAG ACA TTG GAA TAT AAC TTA ACG ACT ACA GAA GTA GTG ATG GAG AAT	1320
	Lys Thr Leu Glu Tyr Asn Leu Thr Thr Thr Glu Val Val Met Glu Asn	
	385 390 395	

- 99 -

	GTA ACA GCC TTC TGG GAG GAG GGA TTT GGG GAA TTA TTT GAG AAA GCA Val Thr Ala Phe Trp Glu Glu Gly Phe Gly Glu Leu Phe Glu Lys Ala 400 405 410	1368
5	AAA CAA AAC AAT AAC AAT AGA AAA ACT TCT AAT GGT GAT GAC AGC CTC Lys Gln Asn Asn Asn Asn Arg Lys Thr Ser Asn Gly Asp Asp Ser Leu 415 420 425	1416
10	TTC TTC AGT AAT TTC TCA CTT CTT GGT ACT CCT GTC CTG AAA GAT ATT Phe Phe Ser Asn Phe Ser Leu Leu Gly Thr Pro Val Leu Lys Asp Ile 430 435 440	1464
15	AAT TTC AAG ATA GAA AGA GGA CAG TTG TTG GCG GTT GCT GGA TCC ACT Asn Phe Lys Ile Glu Arg Gly Gln Leu Leu Ala Val Ala Gly Ser Thr 445 450 455 460	1512
20	GGA GCA GGC AAG ACT TCA CTT CTA ATG ATG ATT ATG GGA GAA CTG GAG Gly Ala Gly Lys Thr Ser Leu Leu Met Met Ile Met Gly Glu Leu Glu 465 470 475	1560
25	CCT TCA GAG GGT AAA ATT AAG CAC AGT GGA AGA ATT TCA TTC TGT TCT Pro Ser Glu Gly Lys Ile Lys His Ser Gly Arg Ile Ser Phe Cys Ser 480 485 490	1608
	CAG TTT TCC TGG ATT ATG CCT GGC ACC ATT AAA GAA AAT ATC ATC TTT Gln Phe Ser Trp Ile Met Pro Gly Thr Ile Lys Glu Asn Ile Ile Phe 495 500 505	1656
30	GGT GTT TCC TAT GAT GAA TAT AGA TAC AGA AGC GTC ATC AAA GCA TGC Gly Val Ser Tyr Asp Glu Tyr Arg Tyr Arg Ser Val Ile Lys Ala Cys 510 515 520	1704
35	CAA CTA GAA GAG GAC ATC TCC AAG TTT GCA GAG AAA GAC AAT ATA GTT Gln Leu Glu Glu Asp Ile Ser Lys Phe Ala Glu Lys Asp Asn Ile Val 525 530 535 540	1752
40	CTT GGA GAA GGT GGA ATC ACA CTG AGT GGA GGT CAA CGA GCA AGA ATT Leu Gly Glu Gly Gly Ile Thr Leu Ser Gly Gly Gln Arg Ala Arg Ile 545 550 555	1800
45	TCT TTA GCA AGA GCA GTA TAC AAA GAT GCT GAT TTG TAT TTA TTA GAC Ser Leu Ala Arg Ala Val Tyr Lys Asp Ala Asp Leu Tyr Leu Leu Asp 560 565 570	1848
	TCT CCT TTT GGA TAC CTA GAT GTT TTA ACA GAA AAA GAA ATA TTT GAA Ser Pro Phe Gly Tyr Leu Asp Val Leu Thr Glu Lys Glu Ile Phe Glu 575 580 585	1896
50	AGC TGT GTC TGT AAA CTG ATG GCT AAC AAA ACT AGG ATT TTG GTC ACT Ser Cys Val Cys Lys Leu Met Ala Asn Lys Thr Arg Ile Leu Val Thr 590 595 600	1944
55	TCT AAA ATG GAA CAT TTA AAG AAA GCT GAC AAA ATA TTA ATT TTG CAT Ser Lys Met Glu His Leu Lys Lys Ala Asp Lys Ile Leu Ile Leu His 605 610 615 620	1992

	GAA GGT AGC AGC TAT TTT TAT GGG ACA TTT TCA GAA CTC CAA AAT CTA	2040
	Glu Gly Ser Ser Tyr Phe Tyr Gly Thr Phe Ser Glu Leu Gln Asn Leu	
	625 630 635	
5	CAG CCA GAC TTT AGC TCA AAA CTC ATG GGA TGT GAT TCT TTC GAC CAA	2088
	Gln Pro Asp Phe Ser Ser Lys Leu Met Gly Cys Asp Ser Phe Asp Gln	
	640 645 650	
10	TTT AGT GCA GAA AGA AGA AAT TCA ATC CTA ACT GAG ACC TTA CAC CGT	2136
	Phe Ser Ala Glu Arg Arg Asn Ser Ile Leu Thr Glu Thr Leu His Arg	
	655 660 665	
15	TTC TCA TTA GAA GGA GAT GCT CCT GTC TCC TGG ACA GAA ACA AAA AAA	2184
	Phe Ser Leu Glu Gly Asp Ala Pro Val Ser Trp Thr Glu Thr Lys Lys	
	670 675 680	
20	CAA TCT TTT AAA CAG ACT GGA GAG TTT GGG GAA AAA AGG AAG AAT TCT	2232
	Gln Ser Phe Lys Gln Thr Gly Glu Phe Gly Glu Lys Arg Lys Asn Ser	
	685 690 695 700	
25	ATT CTC AAT CCA ATC AAC TCT ATA CGA AAA TTT TCC ATT GTG CAA AAG	2280
	Ile Leu Asn Pro Ile Asn Ser Ile Arg Lys Phe Ser Ile Val Gln Lys	
	705 710 715	
30	ACT CCC TTA CAA ATG AAT GGC ATC GAA GAG GAT TCT GAT GAG CCT TTA	2328
	Thr Pro Leu Gln Met Asn Gly Ile Glu Glu Asp Ser Asp Glu Pro Leu	
	720 725 730	
35	GAG AGA AGG CTG TCC TTA GTA CCA GAT TCT GAG CAG GGA GAG GCG ATA	2376
	Glu Arg Arg Leu Ser Leu Val Pro Asp Ser Glu Gln Gly Glu Ala Ile	
	735 740 745	
40	CTG CCT CGC ATC AGC GTG ATC AGC ACT GGC CCC ACG CTT CAG GCA CGA	2424
	Leu Pro Arg Ile Ser Val Ile Ser Thr Gly Pro Thr Leu Gln Ala Arg	
	750 755 760	
45	AGG AGG CAG TCT GTC CTG AAC CTG ATG ACA CAC TCA GTT AAC CAA GGT	2472
	Arg Arg Gln Ser Val Leu Asn Leu Met Thr His Ser Val Asn Gln Gly	
	765 770 775 780	
50	CAG AAC ATT CAC CGA AAG ACA ACA GCA TCC ACA CGA AAA GTG TCA CTG	2520
	Gln Asn Ile His Arg Lys Thr Thr Ala Ser Thr Arg Lys Val Ser Leu	
	785 790 795	
55	GCC CCT CAG GCA AAC TTG ACT GAA CTG GAT ATA TAT TCA AGA AGG TTA	2568
	Ala Pro Gln Ala Asn Leu Thr Glu Leu Asp Ile Tyr Ser Arg Arg Leu	
	800 805 810	
60	TCT CAA GAA ACT GGC TTG GAA ATA AGT GAA GAA ATT AAC GAA GAA GAC	2616
	Ser Gln Glu Thr Gly Leu Glu Ile Ser Glu Glu Ile Asn Glu Glu Asp	
	815 820 825	
65	TTA AAG GAG TGC CTT TTT GAT GAT ATG GAG AGC ATA CCA GCA GTG ACT	2664
	Leu Lys Glu Cys Leu Phe Asp Asp Met Glu Ser Ile Pro Ala Val Thr	
	830 835 840	

- 101 -

5	ACA TGG AAC ACA TAC CTT CGA TAT ATT ACT GTC CAC AAG AGC TTA ATT Thr Trp Asn Thr Tyr Leu Arg Tyr Ile Thr Val His Lys Ser Leu Ile 845 850 855 860	2712
10	TTT GTG CTA ATT TGG TGC TTA GTA ATT TTT CTG GCA GAG GTG GCT GCT Phe Val Leu Ile Trp Cys Leu Val Ile Phe Leu Ala Glu Val Ala Ala 865 870 875	2760
15	TCT TTG GTT GTG CTG TGG CTC CTT GGA AAC ACT CCT CTT CAA GAC AAA Ser Leu Val Val Leu Trp Leu Leu Gly Asn Thr Pro Leu Gln Asp Lys 880 885 890	2808
20	GGG AAT AGT ACT CAT AGT AGA AAT AAC AGC TAT GCA GTG ATT ATC ACC Gly Asn Ser Thr His Ser Arg Asn Asn Ser Tyr Ala Val Ile Ile Thr 895 900 905	2856
25	AGC ACC AGT TCG TAT TAT GTG TTT TAC ATT TAC GTG GGA GTA GCC GAC Ser Thr Ser Ser Tyr Tyr Val Phe Tyr Ile Tyr Val Gly Val Ala Asp 910 915 920	2904
30	ACT TTG CTT GCT ATG GGA TTC TTC AGA GGT CTA CCA CTG GTG CAT ACT Thr Leu Leu Ala Met Gly Phe Phe Arg Gly Leu Pro Leu Val His Thr 925 930 935 940	2952
35	CTA ATC ACA GTG TCG AAA ATT TTA CAC CAC AAA ATG TTA CAT TCT GTT Leu Ile Thr Val Ser Lys Ile Leu His His Lys Met Leu His Ser Val 945 950 955	3000
40	CTT CAA GCA CCT ATG TCA ACC CTC AAC ACG TTG AAA GCA GGT GGG ATT Leu Gln Ala Pro Met Ser Thr Leu Asn Thr Leu Lys Ala Gly Gly Ile 960 965 970	3048
45	CTT AAT AGA TTC TCC AAA GAT ATA GCA ATT TTG GAT GAC CTT CTG CCT Leu Asn Arg Phe Ser Lys Asp Ile Ala Ile Leu Asp Asp Leu Leu Pro 975 980 985	3096
50	CTT ACC ATA TTT GAC TTC ATC CAG TTG TTA TTA ATT GTG ATT GGA GCT Leu Thr Ile Phe Asp Phe Ile Gln Leu Leu Leu Ile Val Ile Gly Ala 990 995 1000	3144
55	ATA GCA GTT GTC GCA GTT TTA CAA CCC TAC ATC TTT GTT GCA ACA GTG Ile Ala Val Val Ala Val Leu Gln Pro Tyr Ile Phe Val Ala Thr Val 1005 1010 1015 1020	3192
60	CCA GTG ATA GTG GCT TTT ATT ATG TTG AGA GCA TAT TTC CTC CAA ACC Pro Val Ile Val Ala Phe Ile Met Leu Arg Ala Tyr Phe Leu Gln Thr 1025 1030 1035	3240
65	TCA CAG CAA CTC AAA CAA CTG GAA TCT GAA GGC AGG AGT CCA ATT TTC Ser Gln Gln Leu Lys Gln Leu Glu Ser Glu Gly Arg Ser Pro Ile Phe 1040 1045 1050	3288
70	ACT CAT CTT GTT ACA AGC TTA AAA GGA CTA TGG ACA CTT CGT GCC TTC Thr His Leu Val Thr Ser Leu Lys Gly Leu Trp Thr Leu Arg Ala Phe 1055 1060 1065	3336



- 102 -

	GGA CGG CAG CCT TAC TTT GAA ACT CTG TTC CAC AAA GCT CTG AAT TTA	3384
	Gly Arg Gln Pro Tyr Phe Glu Thr Leu Phe His Lys Ala Leu Asn Leu	
	1070 1075 1080	
5	CAT ACT GCC AAC TGG TTC TTG TAC CTG TCA ACA CTG CGC TGG TTC CAA	3432
	His Thr Ala Asn Trp Phe Leu Tyr Leu Ser Thr Leu Arg Trp Phe Gln	
	1085 1090 1095 1100	
10	ATG AGA ATA GAA ATG ATT TTT GTC ATC TTC TTC ATT GCT GTT ACC TTC	3480
	Met Arg Ile Glu Met Ile Phe Val Ile Phe Phe Ile Ala Val Thr Phe	
	1105 1110 1115	
15	ATT TCC ATT TTA ACA ACA GGA GAA GGA GAA GGA AGA GTT GGT ATT ATC	3528
	Ile Ser Ile Leu Thr Thr Gly Glu Gly Glu Gly Arg Val Gly Ile Ile	
	1120 1125 1130	
20	CTG ACT TTA GCC ATG AAT ATC ATG AGT ACA TTG CAG TGG GCT GTA AAC	3576
	Leu Thr Leu Ala Met Asn Ile Met Ser Thr Leu Gln Trp Ala Val Asn	
	1135 1140 1145	
25	TCC AGC ATA GAT GTG GAT AGC TTG ATG CGA TCT GTG AGC CGA GTC TTT	3624
	Ser Ser Ile Asp Val Asp Ser Leu Met Arg Ser Val Ser Arg Val Phe	
	1150 1155 1160	
30	AAG TTC ATT GAC ATG CCA ACA GAA GGT AAA CCT ACC AAG TCA ACC AAA	3672
	Lys Phe Ile Asp Met Pro Thr Glu Gly Lys Pro Thr Lys Ser Thr Lys	
	1165 1170 1175 1180	
35	CCA TAC AAG AAT GGC CAA CTC TCG AAA GTT ATG ATT ATT GAG AAT TCA	3720
	Pro Tyr Lys Asn Gly Gln Leu Ser Lys Val Met Ile Ile Glu Asn Ser	
	1185 1190 1195	
40	CAC GTG AAG AAA GAT GAC ATC TGG CCC TCA GGG GGC CAA ATG ACT GTC	3768
	His Val Lys Lys Asp Asp Ile Trp Pro Ser Gly Gly Gln Met Thr Val	
	1200 1205 1210	
45	AAA GAT CTC ACA GCA AAA TAC ACA GAA GGT GGA AAT GCC ATA TTA GAG	3816
	Lys Asp Leu Thr Ala Lys Tyr Thr Glu Gly Gly Asn Ala Ile Leu Glu	
	1215 1220 1225	
50	AAC ATT TCC TTC TCA ATA AGT CCT GGC CAG AGG GTG GGC CTC TTG GGA	3864
	Asn Ile Ser Phe Ser Ile Ser Pro Gly Gln Arg Val Gly Leu Leu Gly	
	1230 1235 1240	
55	AGA ACT GGA TCA GGG AAG AGT ACT TTG TTA TCA GCT TTT TTG AGA CTA	3912
	Arg Thr Gly Ser Gly Lys Ser Thr Leu Leu Ser Ala Phe Leu Arg Leu	
	1245 1250 1255 1260	
50	CTG AAC ACT GAA GGA GAA ATC CAG ATC GAT GGT GTG TCT TGG GAT TCA	3960
	Leu Asn Thr Glu Gly Glu Ile Gln Ile Asp Gly Val Ser Trp Asp Ser	
	1265 1270 1275	
55	ATA ACT TTG CAA CAG TGG AGG AAA GCC TTT GGA GTG ATA CCA CAG AAA	4008
	Ile Thr Leu Gln Gln Trp Arg Lys Ala Phe Gly Val Ile Pro Gln Lys	
	1280 1285 1290	

5	GTA TTT ATT TTT TCT GGA ACA TTT AGA AAA AAC TTG GAT CCC TAT GAA Val Phe Ile Phe Ser Gly Thr Phe Arg Lys Asn Leu Asp Pro Tyr Glu 1295 1300 1305	4056
	CAG TGG AGT GAT CAA GAA ATA TGG AAA GTT GCA GAT GAG GTT GGG CTC Gln Trp Ser Asp Gln Glu Ile Trp Lys Val Ala Asp Glu Val Gly Leu 1310 1315 1320	4104
10	AGA TCT GTG ATA GAA CAG TTT CCT GGG AAG CTT GAC TTT GTC CTT GTG Arg Ser Val Ile Glu Gln Phe Pro Gly Lys Leu Asp Phe Val Leu Val 1325 1330 1335 1340	4152
15	GAT GGG GGC TGT GTC CTA AGC CAT GGC CAC AAG CAG TTG ATG TGC TTG Asp Gly Gly Cys Val Leu Ser His Gly His Lys Gln Leu Met Cys Leu 1345 1350 1355	4200
20	GCT AGA TCT GTT CTC AGT AAG GCG AAG ATC TTG CTG CTT GAT GAA CCC Ala Arg Ser Val Leu Ser Lys Ala Lys Ile Leu Leu Leu Asp Glu Pro 1360 1365 1370	4248
25	AGT GCT CAT TTG GAT CCA GTA ACA TAC CAA ATA ATT AGA AGA ACT CTA Ser Ala His Leu Asp Pro Val Thr Tyr Gln Ile Ile Arg Arg Thr Leu 1375 1380 1385	4296
	AAA CAA GCA TTT GCT GAT TGC ACA GTA ATT CTC TGT GAA CAC AGG ATA Lys Gln Ala Phe Ala Asp Cys Thr Val Ile Leu Cys Glu His Arg Ile 1390 1395 1400	4344
30	GAA GCA ATG CTG GAA TGC CAA CAA TTT TTG GTC ATA GAA GAG AAC AAA Glu Ala Met Leu Glu Cys Gln Gln Phe Leu Val Ile Glu Glu Asn Lys 1405 1410 1415 1420	4392
35	GTG CGG CAG TAC GAT TCC ATC CAG AAA CTG CTG AAC GAG AGG AGC CTC Val Arg Gln Tyr Asp Ser Ile Gln Lys Leu Leu Asn Glu Arg Ser Leu 1425 1430 1435	4440
40	TTC CGG CAA GCC ATC AGC CCC TCC GAC AGG GTG AAG CTC TTT CCC CAC Phe Arg Gln Ala Ile Ser Pro Ser Asp Arg Val Lys Leu Phe Pro His 1440 1445 1450	4488
45	CGG AAC TCA AGC AAG TGC AAG TCT AAG CCC CAG ATT GCT GCT CTG AAA Arg Asn Ser Ser Lys Cys Lys Ser Lys Pro Gln Ile Ala Ala Leu Lys 1455 1460 1465	4536
	GAG GAG ACA GAA GAA GAG GTG CAA GAT ACA AGG CTT TAGAGAGCAG Glu Glu Thr Glu Glu Glu Val Gln Asp Thr Arg Leu 1470 1475 1480	4582
50	CATAAATGTT GACATGGGAC ATTTGCTCAT GGAATTGGAG CTCGTGGGAC AGTCACCTCA TGGAATTGGA GCTCGTGGAA CAGTTACCTC TGCCTCAGAA AACAAGGATG AATTAAGTTT	4642 4702
55	TTTTTTAAAA AAGAAACATT TGGTAAGGGG AATTGAGGAC ACTGATATGG GTCTTGATAA ATGGCTTCCT GGCAATAGTC AAATTGTGTG AAAGGTACTT CAAATCCTTG AAGATTTACC	4762 4822
	ACTTGTGTTT TGCAAGCCAG ATTTTCCTGA AAACCCTTGC CATGTGCTAG TAATTGGAAA	4882

- 104 -

GGCAGCTCTA AATGTCAATC AGCCTAGTTG ATCAGCTTAT TGTCTAGTGA AACTCGTTAA 4942  
 TTTGTTAGTGT TGGAGAAGAA CTGAAATCAT ACTTCTTAGG GTTATGATTA AGTAATGATA 5002  
 5 ACTGGAAACT TCAGCGGTTT ATATAAGCTT GTATTCCCTT TTCTCTCCTC TCCCCATGAT 5062  
 GTTTAGAAAC ACAACTATAT TGTTTGCTAA GCATTCCAAC TATCTCATTT CCAAGCAAGT 5122  
 ATTAGAATAC CACAGGAACC ACAAGACTGC ACATCAAAAT ATGCCCCATT CAACATCTAG 5182  
 10 TGAGCAGTCA GGAAAGAGAA CTTCCAGATC CTGGAATCA GGGTTAGTAT TGTCCAGGTC 5242  
 TACCAAAAAT CTCAATATTT CAGATAATCA CAATACATCC CTTACCTGGG AAAGGGCTGT 5302  
 15 TATAATCTTT CACAGGGGAC AGGATGGTTC CCTTGATGAA GAAGTTGATA TGCCTTTTCC 5362  
 CAACTCCAGA AAGTGACAAG CTCACAGACC TTTGAACTAG AGTTTAGCTG GAAAAGTATG 5422  
 TTAGTGCAAA TTGTCACAGG ACAGCCCTTC TTTCCACAGA AGCTCCAGGT AGAGGGTGTG 5482  
 20 TAAGTAGATA GGCCATGGGC ACTGTGGGTA GACACACATG AAGTCCAAGC ATTTAGATGT 5542  
 ATAGGTTGAT GGTGGTATGT TTTCAGGCTA GATGTATGTA CTTCATGCTG TCTACACTAA 5602  
 25 GAGAGAATGA GAGACACACT GAAGAAGCAC CAATCATGAA TTAGTTTTAT ATGCTTCTGT 5662  
 TTTATAATTT TGTGAAGCAA AATTTTTTCT CTAGGAAATA TTTATTTTAA TAATGTTTCA 5722  
 AACATATATT ACAATGCTGT ATTTTAAAG AATGATTATG AATTACATTT GTATAAAATA 5782  
 30 ATTTTATAT TTGAAATATT GACTTTTTAT GGCACTAGTA TTTTATGAA ATATTATGTT 5842  
 AAAACTGGGA CAGGGGAGAA CCTAGGGTGA TATTAAACCAG GGGCCATGAA TCACCTTTTG 5902  
 35 GTCTGGAGGG AAGCCTTGGG GCTGATCGAG TTGTTGCCCA CAGCTGTATG ATTCCCAGCC 5962  
 AGACACAGCC TCTTAGATGC AGTTCTGAAG AAGATGGTAC CACCAGTCTG ACTGTTTCCA 6022  
 TCAAGGGTAC ACTGCCTTCT CAACTCCAAA CTGACTCTTA AGAAGACTGC ATTATATTTA 6082  
 40 TTACTGTAAG AAAATATCAC TTGTCAATAA AATCCATACA TTTGTGT 6129

## (2) INFORMATION FOR SEQ ID NO:2:

45

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1480 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

50

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

55 Met Gln Arg Ser Pro Leu Glu Lys Ala Ser Val Val Ser Lys Leu Phe  
 1 5 10 15

	Phe	Ser	Trp	Thr	Arg	Pro	Ile	Leu	Arg	Lys	Gly	Tyr	Arg	Gln	Arg	Leu	
				20					25					30			
5	Glu	Leu	Ser	Asp	Ile	Tyr	Gln	Ile	Pro	Ser	Val	Asp	Ser	Ala	Asp	Asn	
		35					40					45					
	Leu	Ser	Glu	Lys	Leu	Glu	Arg	Glu	Trp	Asp	Arg	Glu	Leu	Ala	Ser	Lys	
		50				55					60						
10	Lys	Asn	Pro	Lys	Leu	Ile	Asn	Ala	Leu	Arg	Arg	Cys	Phe	Phe	Trp	Arg	
	65				70					75						80	
	Phe	Met	Phe	Tyr	Gly	Ile	Phe	Leu	Tyr	Leu	Gly	Glu	Val	Thr	Lys	Ala	
15					85				90						95		
	Val	Gln	Pro	Leu	Leu	Leu	Gly	Arg	Ile	Ile	Ala	Ser	Tyr	Asp	Pro	Asp	
			100					105						110			
20	Asn	Lys	Glu	Glu	Arg	Ser	Ile	Ala	Ile	Tyr	Leu	Gly	Ile	Gly	Leu	Cys	
		115					120					125					
	Leu	Leu	Phe	Ile	Val	Arg	Thr	Leu	Leu	Leu	His	Pro	Ala	Ile	Phe	Gly	
	130					135					140						
25	Leu	His	His	Ile	Gly	Met	Gln	Met	Arg	Ile	Ala	Met	Phe	Ser	Leu	Ile	
	145				150				155							160	
	Tyr	Lys	Lys	Thr	Leu	Lys	Leu	Ser	Ser	Arg	Val	Leu	Asp	Lys	Ile	Ser	
30				165					170					175			
	Ile	Gly	Gln	Leu	Val	Ser	Leu	Leu	Ser	Asn	Asn	Leu	Asn	Lys	Phe	Asp	
			180					185					190				
35	Glu	Gly	Leu	Ala	Leu	Ala	His	Phe	Val	Trp	Ile	Ala	Pro	Leu	Gln	Val	
		195					200					205					
	Ala	Leu	Leu	Met	Gly	Leu	Ile	Trp	Glu	Leu	Leu	Gln	Ala	Ser	Ala	Phe	
40		210				215					220						
	Cys	Gly	Leu	Gly	Phe	Leu	Ile	Val	Leu	Ala	Leu	Phe	Gln	Ala	Gly	Leu	
	225				230					235					240		
	Gly	Arg	Met	Met	Met	Lys	Tyr	Arg	Asp	Gln	Arg	Ala	Gly	Lys	Ile	Ser	
45				245					250					255			
	Glu	Arg	Leu	Val	Ile	Thr	Ser	Glu	Met	Ile	Glu	Asn	Ile	Gln	Ser	Val	
		260						265					270				
50	Lys	Ala	Tyr	Cys	Trp	Glu	Glu	Ala	Met	Glu	Lys	Met	Ile	Glu	Asn	Leu	
		275					280					285					
	Arg	Gln	Thr	Glu	Leu	Lys	Leu	Thr	Arg	Lys	Ala	Ala	Tyr	Val	Arg	Tyr	
		290				295					300						
55	Phe	Asn	Ser	Ser	Ala	Phe	Phe	Phe	Ser	Gly	Phe	Phe	Val	Val	Phe	Leu	
	305				310					315						320	

- 106 -

Ser Val Leu Pro Tyr Ala Leu Ile Lys Gly Ile Ile Leu Arg Lys Ile  
 325 330 335

5 Phe Thr Thr Ile Ser Phe Cys Ile Val Leu Arg Met Ala Val Thr Arg  
 340 345 350

Gln Phe Pro Trp Ala Val Gln Thr Trp Tyr Asp Ser Leu Gly Ala Ile  
 355 360 365

10 Asn Lys Ile Gln Asp Phe Leu Gln Lys Gln Glu Tyr Lys Thr Leu Glu  
 370 375 380

Tyr Asn Leu Thr Thr Thr Glu Val Val Met Glu Asn Val Thr Ala Phe  
 15 385 390 395 400

Trp Glu Glu Gly Phe Gly Glu Leu Phe Glu Lys Ala Lys Gln Asn Asn  
 405 410 415

20 Asn Asn Arg Lys Thr Ser Asn Gly Asp Asp Ser Leu Phe Phe Ser Asn  
 420 425 430

Phe Ser Leu Leu Gly Thr Pro Val Leu Lys Asp Ile Asn Phe Lys Ile  
 435 440 445

25 Glu Arg Gly Gln Leu Leu Ala Val Ala Gly Ser Thr Gly Ala Gly Lys  
 450 455 460

Thr Ser Leu Leu Met Met Ile Met Gly Glu Leu Glu Pro Ser Glu Gly  
 30 465 470 475 480

Lys Ile Lys His Ser Gly Arg Ile Ser Phe Cys Ser Gln Phe Ser Trp  
 485 490 495

35 Ile Met Pro Gly Thr Ile Lys Glu Asn Ile Ile Phe Gly Val Ser Tyr  
 500 505 510

Asp Glu Tyr Arg Tyr Arg Ser Val Ile Lys Ala Cys Gln Leu Glu Glu  
 40 515 520 525

Asp Ile Ser Lys Phe Ala Glu Lys Asp Asn Ile Val Leu Gly Glu Gly  
 530 535 540

Gly Ile Thr Leu Ser Gly Gly Gln Arg Ala Arg Ile Ser Leu Ala Arg  
 45 545 550 555 560

Ala Val Tyr Lys Asp Ala Asp Leu Tyr Leu Leu Asp Ser Pro Phe Gly  
 565 570 575

50 Tyr Leu Asp Val Leu Thr Glu Lys Glu Ile Phe Glu Ser Cys Val Cys  
 580 585 590

Lys Leu Met Ala Asn Lys Thr Arg Ile Leu Val Thr Ser Lys Met Glu  
 595 600 605

55 His Leu Lys Lys Ala Asp Lys Ile Leu Ile Leu His Glu Gly Ser Ser  
 610 615 620

- 107 -

Tyr Phe Tyr Gly Thr Phe Ser Glu Leu Gln Asn Leu Gln Pro Asp Phe  
 625 630 635 640  
 5 Ser Ser Lys Leu Met Gly Cys Asp Ser Phe Asp Gln Phe Ser Ala Glu  
 645 650 655  
 Arg Arg Asn Ser Ile Leu Thr Glu Thr Leu His Arg Phe Ser Leu Glu  
 660 665 670  
 10 Gly Asp Ala Pro Val Ser Trp Thr Glu Thr Lys Lys Gln Ser Phe Lys  
 675 680 685  
 Gln Thr Gly Glu Phe Gly Glu Lys Arg Lys Asn Ser Ile Leu Asn Pro  
 690 695 700  
 15 Ile Asn Ser Ile Arg Lys Phe Ser Ile Val Gln Lys Thr Pro Leu Gln  
 705 710 715 720  
 Met Asn Gly Ile Glu Glu Asp Ser Asp Glu Pro Leu Glu Arg Arg Leu  
 725 730 735  
 Ser Leu Val Pro Asp Ser Glu Gln Gly Glu Ala Ile Leu Pro Arg Ile  
 740 745 750  
 25 Ser Val Ile Ser Thr Gly Pro Thr Leu Gln Ala Arg Arg Arg Gln Ser  
 755 760 765  
 Val Leu Asn Leu Met Thr His Ser Val Asn Gln Gly Gln Asn Ile His  
 770 775 780  
 30 Arg Lys Thr Thr Ala Ser Thr Arg Lys Val Ser Leu Ala Pro Gln Ala  
 785 790 795 800  
 Asn Leu Thr Glu Leu Asp Ile Tyr Ser Arg Arg Leu Ser Gln Glu Thr  
 805 810 815  
 Gly Leu Glu Ile Ser Glu Glu Ile Asn Glu Glu Asp Leu Lys Glu Cys  
 820 825 830  
 40 Leu Phe Asp Asp Met Glu Ser Ile Pro Ala Val Thr Thr Trp Asn Thr  
 835 840 845  
 Tyr Leu Arg Tyr Ile Thr Val His Lys Ser Leu Ile Phe Val Leu Ile  
 850 855 860  
 Trp Cys Leu Val Ile Phe Leu Ala Glu Val Ala Ala Ser Leu Val Val  
 865 870 875 880  
 50 Leu Trp Leu Leu Gly Asn Thr Pro Leu Gln Asp Lys Gly Asn Ser Thr  
 885 890 895  
 His Ser Arg Asn Asn Ser Tyr Ala Val Ile Ile Thr Ser Thr Ser Ser  
 900 905 910  
 55 Tyr Tyr Val Phe Tyr Ile Tyr Val Gly Val Ala Asp Thr Leu Leu Ala  
 915 920 925

- 108 -

Met Gly Phe Phe Arg Gly Leu Pro Leu Val His Thr Leu Ile Thr Val  
930 935 940

5 Ser Lys Ile Leu His His Lys Met Leu His Ser Val Leu Gln Ala Pro  
945 950 955 960

Met Ser Thr Leu Asn Thr Leu Lys Ala Gly Gly Ile Leu Asn Arg Phe  
965 970 975

10 Ser Lys Asp Ile Ala Ile Leu Asp Asp Leu Leu Pro Leu Thr Ile Phe  
980 985 990

15 Asp Phe Ile Gln Leu Leu Leu Ile Val Ile Gly Ala Ile Ala Val Val  
995 1000 1005

Ala Val Leu Gln Pro Tyr Ile Phe Val Ala Thr Val Pro Val Ile Val  
1010 1015 1020

20 Ala Phe Ile Met Leu Arg Ala Tyr Phe Leu Gln Thr Ser Gln Gln Leu  
1025 1030 1035 1040

Lys Gln Leu Glu Ser Glu Gly Arg Ser Pro Ile Phe Thr His Leu Val  
1045 1050 1055

25 Thr Ser Leu Lys Gly Leu Trp Thr Leu Arg Ala Phe Gly Arg Gln Pro  
1060 1065 1070

30 Tyr Phe Glu Thr Leu Phe His Lys Ala Leu Asn Leu His Thr Ala Asn  
1075 1080 1085

Trp Phe Leu Tyr Leu Ser Thr Leu Arg Trp Phe Gln Met Arg Ile Glu  
1090 1095 1100

35 Met Ile Phe Val Ile Phe Phe Ile Ala Val Thr Phe Ile Ser Ile Leu  
1105 1110 1115 1120

Thr Thr Gly Glu Gly Glu Gly Arg Val Gly Ile Ile Leu Thr Leu Ala  
1125 1130 1135

40 Met Asn Ile Met Ser Thr Leu Gln Trp Ala Val Asn Ser Ser Ile Asp  
1140 1145 1150

Val Asp Ser Leu Met Arg Ser Val Ser Arg Val Phe Lys Phe Ile Asp  
1155 1160 1165

Met Pro Thr Glu Gly Lys Pro Thr Lys Ser Thr Lys Pro Tyr Lys Asn  
1170 1175 1180

50 Gly Gln Leu Ser Lys Val Met Ile Ile Glu Asn Ser His Val Lys Lys  
1185 1190 1195 1200

Asp Asp Ile Trp Pro Ser Gly Gly Gln Met Thr Val Lys Asp Leu Thr  
1205 1210 1215

55 Ala Lys Tyr Thr Glu Gly Gly Asn Ala Ile Leu Glu Asn Ile Ser Phe  
1220 1225 1230

- 109 -

- Ser Ile Ser Pro Gly Gln Arg Val Gly Leu Leu Gly Arg Thr Gly Ser  
1235 1240 1245
- 5 Gly Lys Ser Thr Leu Leu Ser Ala Phe Leu Arg Leu Leu Asn Thr Glu  
1250 1255 1260
- Gly Glu Ile Gln Ile Asp Gly Val Ser Trp Asp Ser Ile Thr Leu Gln  
1265 1270 1275 1280
- 10 Gln Trp Arg Lys Ala Phe Gly Val Ile Pro Gln Lys Val Phe Ile Phe  
1285 1290 1295
- Ser Gly Thr Phe Arg Lys Asn Leu Asp Pro Tyr Glu Gln Trp Ser Asp  
15 1300 1305 1310
- Gln Glu Ile Trp Lys Val Ala Asp Glu Val Gly Leu Arg Ser Val Ile  
1315 1320 1325
- 20 Glu Gln Phe Pro Gly Lys Leu Asp Phe Val Leu Val Asp Gly Gly Cys  
1330 1335 1340
- Val Leu Ser His Gly His Lys Gln Leu Met Cys Leu Ala Arg Ser Val  
1345 1350 1355 1360
- 25 Leu Ser Lys Ala Lys Ile Leu Leu Leu Asp Glu Pro Ser Ala His Leu  
1365 1370 1375
- Asp Pro Val Thr Tyr Gln Ile Ile Arg Arg Thr Leu Lys Gln Ala Phe  
30 1380 1385 1390
- Ala Asp Cys Thr Val Ile Leu Cys Glu His Arg Ile Glu Ala Met Leu  
1395 1400 1405
- 35 Glu Cys Gln Gln Phe Leu Val Ile Glu Glu Asn Lys Val Arg Gln Tyr  
1410 1415 1420
- Asp Ser Ile Gln Lys Leu Leu Asn Glu Arg Ser Leu Phe Arg Gln Ala  
1425 1430 1435 1440
- 40 Ile Ser Pro Ser Asp Arg Val Lys Leu Phe Pro His Arg Asn Ser Ser  
1445 1450 1455
- Lys Cys Lys Ser Lys Pro Gln Ile Ala Ala Leu Lys Glu Glu Thr Glu  
45 1460 1465 1470
- Glu Glu Val Gln Asp Thr Arg Leu  
1475 1480
- 50 (2) INFORMATION FOR SEQ ID NO:3:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5635 base pairs
- (B) TYPE: nucleic acid
- 55 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

5	CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT	60
	TTGTGACGTG GCGCGGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCGGAAGTGT	120
	GATGTTGCAA GTGTGGCGGA ACACATGTAA GCGCCGGATG TGGTAAAAGT GACGTTTTTG	180
10	GTGTGCGCCG GTGTATACGG GAAGTGACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG	240
	TAAATTTGGG CGTAACCAAG TAATGTTTGG CCATTTTCGC GGGAAAAGT AATAAGAGGA	300
15	AGTGAAATCT GAATAATTCT GTGTTACTCA TAGCGCGTAA TATTTGTCTA GGGCCGCGGG	360
	GACTTTGACC GTTTACGTGG AGACTCGCCC AGGTGTTTTT CTCAGGTGTT TTCCGCGTTC	420
	CGGGTCAAAG TTGGCGTTTT ATTATTATAG TCAGCTGACG CGCAGTGTAT TTATACCCGG	480
20	TGAGTTCCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTTCTCC TCCGAGCCGC	540
	TCCGAGCTAG TAACGGCCGC CAGTGTGCTG CAGATATCAA AGTCGACGGT ACCCGAGAGA	600
25	CCATGCAGAG GTCGCCTCTG GAAAAGGCCA GCGTTGTCTC CAAACTTTTT TTCAGCTGGA	660
	CCAGACCAAT TTTGAGGAAA GGATACAGAC AGCGCCTGGA ATTGTCAGAC ATATACCAAA	720
	TCCCTTCTGT TGATTCTGCT GACAATCTAT CTGAAAAATT GGAAAGAGAA TGGGATAGAG	780
30	AGCTGGCTTC AAAGAAAAAT CCTAAACTCA TTAATGCCCT TCGGCGATGT TTTTCTGGA	840
	GATTATGTT CTATGGAATC TTTTATATT TAGGGGAAGT CACCRAAGCA GTACAGCCTC	900
35	TCCTACTGGG AAGAATCATA GCTTCCTATG ACCCGGATAA CAAGGAGGAA CGCTCTATCG	960
	CGATTTATCT AGGCATAGGC TTATGCCTTC TCTTTATTGT GAGGACACTG CTCCTACACC	1020
	CAGCCATTTT TGGCCTTCAT CACATTGGAA TGCAGATGAG AATAGCTATG TTTAGTTTGA	1080
40	TTTATAAGAA GACTTTAAAG CTGTCAAGCC GTGTTCTAGA TAAAATAAGT ATTGGACAAC	1140
	TTGTTAGTCT CCTTTCCAAC AACCTGAACA AATTGATGA AGGACTTGCA TTGGCACATT	1200
45	TCGTGTGGAT CGCTCCTTTG CAAGTGGCAC TCCTCATGGG GCTAATCTGG GAGTTGTTAC	1260
	AGGCGTCTGC CTTCTGTGGA CTTGGTTTCC TGATAGTCCT TGCCCTTTTT CAGGCTGGGC	1320
	TAGGGAGAAT GATGATGAAG TACAGAGATC AGAGAGCTGG GAAGATCAGT GAAAGACTTG	1380
50	TGATTACCTC AGAAATGATT GAAAACATCC AATCTGTTAA GGCATACTGC TGGGAAGAAG	1440
	CAATGAAAAA AATGATTGAA AACTTAAGAC AAACAGAACT GAAACTGACT CGGAAGGCAG	1500
55	CCTATGTGAG AACTTCAAT AGCTCAGCCT TCTTCTTCTC AGGGTTCTTT GTGGTGTTTT	1560
	TATCTGTGCT TCCCTATGCA CTAATCAAAG GAATCATCCT CCGGAAAATA TTCACCACCA	1620
	TCTCATTCTG CATGTTCTG CGCATGGCGG TCACTCGGCA ATTCCCTGG GCTGTACAAA	1680

	CATGGTATGA CTCTCTTGGG GCAATAAACA AAATACAGGA TTTCTTACAA AAGCAAGAAT	1740
	ATAAGACATT GGAATATAAC TTAACGACTA CAGAAGTAGT GATGGAGAAT GTAACAGCCT	1800
5	TCTGGGAGGA GGGATTTGGG GAATTATTTG AGAAAGCAAA ACAAACAAT AACAAATAGAA	1860
	AAACTTCTAA TGGTGATGAC AGCCTCTTCT TCAGTAATTT CTCACTTCTT GGTACTCCTG	1920
10	TCCTGAAAGA TATTAATTTC AAGATAGAAA GAGGACAGTT GTTGGCGGTT GCTGGATCCA	1980
	CTGGAGCAGG CAAGACTTCA CTTCTAATGA TGATTATGGG AGAACTGGAG CCTTCAGAGG	2040
	GTAAAATTAA GCACAGTGA AGAATTTTCAT TCTGTTCTCA GTTTTCCTGG ATTATGCCTG	2100
15	GCACCATTAA AGAAAATATC ATCTTTGGTG TTTCTATGA TGAATATAGA TACAGAAGCG	2160
	TCATCAAAGC ATGCCAACTA GAAGAGGACA TCTCCAAGTT TGCAGAGAAA GACAATATAG	2220
20	TTCTTGGAGA AGGTGGAATC AACTGAGTG GAGGTCAACG AGCAAGAATT TCTTTAGCAA	2280
	GAGCAGTATA CAAAGATGCT GATTTGTATT TATTAGACTC TCCTTTTGA TACCTAGATG	2340
	TTTTAACAGA AAAAGAAATA TTTGAAAGCT GTGTCTGTAA ACTGATGGCT AACAAAATA	2400
25	GGATTTTGGT CACTTCTAAA ATGGAACATT TAAAGAAAGC TGACAAAATA TTAATTTTGC	2460
	ATGAAGGTAG CAGCTATTTT TATGGGACAT TTTCAGAACT CCAAAATCTA CAGCCAGACT	2520
30	TTAGCTCAAA ACTCATGGGA TGTGATTCTT TCGACCAATT TAGTGCAGAA AGAAGAAATT	2580
	CAATCCTAAC TGAGACCTTA CACCGTTTCT CATTAGAAGG AGATGCTCCT GTCTCCTGGA	2640
	CAGAAACAAA AAAACAATCT TTTAAACAGA CTGGAGAGTT TGGGGAAAAA AGGAAGAATT	2700
35	CTATTCTCAA TCCAATCAAC TCTATACGAA AATTTTCCAT TGTGCAAAAG ACTCCCTTAC	2760
	AAATGAATGG CATCGAAGAG GATTCTGATG AGCCTTTAGA GAGAAGGCTG TCCTTAGTAC	2820
40	CAGATTCTGA GCAGGGAGAG GCGATACTGC CTCGCATCAG CGTGATCAGC ACTGGCCCCA	2880
	CGCTTCAGGC ACGAAGGAGG CAGTCTGTCC TGAACCTGAT GACACACTCA GTTAACCAAG	2940
	GTCAGAACAT TCACCGAAAG ACAACAGCAT CCACACGAAA AGTGTCACCTG GCCCCTCAGG	3000
45	CAAACCTGAC TGAACCTGGAT ATATATTCAA GAAGGTTATC TCAAGAACT GGCTTGGAAA	3060
	TAAGTGAAGA AATTAACGAA GAAGACTTAA AGGAGTGCCT TTTTGATGAT ATGGAGAGCA	3120
50	TACCAGCAGT GACTACATGG AACACATACC TTCGATATAT TACTGTCCAC AAGAGCTTAA	3180
	TTTTTGTGCT AATTTGGTGC TTAGTAATTT TTCTGGCAGA GGTGGCTGCT TCTTTGGTTG	3240
	TGCTGTGGCT CCTTGGAAC ACTCCTCTTC AAGACAAAGG GAATAGTACT CATAGTAGAA	3300
55	ATAACAGCTA TGCAGTGATT ATCACCAGCA CCAGTTCGTA TTATGTGTTT TACATTACG	3360
	TGGGAGTAGC CGACACTTTG CTTGCTATGG GATTCTTCAG AGGTCTACCA CTGGTGCATA	3420
	CTCTAATCAC AGTGTGAAA ATTTTACACC ACAAATGTT ACATTCTGTT CTTCAAGCAC	3480

	CTATGTCAAC CCTCAACACG TTGAAAGCAG GTGGGATTCT TAATAGATTC TCCAAAGATA	3540
5	TAGCAATTTT GGATGACCTT CTGCCTCTTA CCATATTGA CTTTCATCCAG TTGTTATTAA	3600
	TTGTGATTGG AGCTATAGCA GTTGTGCGCAG TTTTACAACC CTACATCTTT GTTGCAACAG	3660
	TGCCAGTGAT AGTGGCTTTT ATTATGTTGA GAGCATATTT CCTCCAAACC TCACAGCAAC	3720
10	TCAAACAACCT GGAATCTGAA GGCAGGAGTC CAATTTTCAC TCATCTTGTT ACAAGCTTAA	3780
	AAGGACTATG GACACTTCGT GCCTTCGGAC GGCAGCCTTA CTTTGAAACT CTGTTCCACA	3840
15	AAGCTCTGAA TTTACATACT GCCAACTGGT TCTTGTACCT GTCAACACTG CGCTGGTTCC	3900
	AAATGAGAAT AGAAATGATT TTTGTCATCT TCTTCATTGC TGTTACCTTC ATTTCCATT	3960
	TAACAACAGG AGAAGGAGAA GGAAGAGTTG GTATTATCCT GACTTTAGCC ATGAATATCA	4020
20	TGAGTACATT GCAGTGGGCT GTAAACTCCA GCATAGATGT GGATAGCTTG ATGCGATCTG	4080
	TGAGCCGAGT CTTTAAGTTC ATTGACATGC CAACAGAAGG TAAACCTACC AAGTCAACCA	4140
25	AACCATACAA GAATGGCCAA CTCTCGAAAG TTATGATTAT TGAGAATTCA CACGTGAAGA	4200
	AAGATGACAT CTGGCCCTCA GGGGGCCAAA TGAAGTGTCAA AGATCTCACA GCAAATACA	4260
	CAGAAGGTGG AAATGCCATA TTAGAGAACA TTCTCTTCTC AATAAGTCCT GGCCAGAGGG	4320
30	TGGGCCTCTT GGGAAGAACT GGATCAGGGA AGAGTACTTT GTTATCAGCT TTTTGTAGAC	4380
	TACTGAACAC TGAAGGAGAA ATCCAGATCG ATGGTGTGTC TTGGGATTCA ATAACCTTGC	4440
35	AACAGTGGAG GAAAGCCTTT GGAGTGATAC CACAGAAAGT ATTTATTTTT TCTGGAACAT	4500
	TTAGAAAAAA CTTGGATCCC TATGAACAGT GGAGTGATCA AGAAATATGG AAAGTTGCAG	4560
	ATGAGGTTGG GCTCAGATCT GTGATAGAAC AGTTTCCTGG GAAGCTTGAC TTTGTCCTTG	4620
40	TGGATGGGGG CTGTGTCCTA AGCCATGGCC ACAAGCAGTT GATGTGCTTG GCTAGATCTG	4680
	TTCTCAGTAA GGCGAAGATC TTGCTGCTTG ATGAACCCAG TGCTCATTTG GATCCAGTAA	4740
45	CATACCAAAT AATTAGAAGA ACTCTAAAAC AAGCATTGTC TGATTGCACA GTAATTCTCT	4800
	GTGAACACAG GATAGAAGCA ATGCTGGAAT GCCAACAATT TTTGGTCATA GAAGAGAACA	4860
	AAGTGCGGCA GTACGATTCC ATCCAGAAAC TGCTGAACGA GAGGAGCCTC TTCCGGCAAG	4920
50	CCATCAGCCC CTCCGACAGG GTGAAGCTCT TTCCCCACCG GAACTCAAGC AAGTGCAAGT	4980
	CTAAGCCCCA GATTGCTGCT CTGAAAGAGG AGACAGAAGA AGAGGTGCAA GATACAAGGC	5040
55	TTTAGAGAGC AGCATAAATG TTGACATGGG ACATTTGCTC ATGGAATTGG AGGTAGCGGA	5100
	TTGAGGTACT GAAATGTGTG GCGTGGCTT AAGGGTGGGA AAGAATATAT AAGGTGGGGG	5160
	TCTCATGTAG TTTTGTATCT GTTTTGCAGC AGCCGCCGCC ATGAGCGCCA ACTCGTTTGA	5220

- 113 -

TGGAAGCATT GTGAGCTCAT ATTTGACAAC GCGCATGCCC CCATGGGCCG GGGTGCCTCA 5280  
GAATGTGATG GGCTCCAGCA TTGATGGTCG CCCCCTCCTG CCCGCAAACCT CTACTACCTT 5340  
5 GACCTACGAG ACCGTGTCTG GAACGCCGTT GGAGACTGCA GCCTCCGCCG CCGCTTCAGC 5400  
CGCTGCAGCC ACCGCCCGCG GGATTGTGAC TGACTTTGCT TTCCTGAGCC CGCTTGCAAG 5460  
CAGTGCAGCT TCCCGTTTCAT CCGCCCGCGA TGACAAGTTG ACGGCTCTTT TGGCACAATT 5520  
10 GGATTCTTTG ACCCGGGAAC TTAATGTCGT TTCTCAGCAG CTGTTGGATC TCGCCAGCA 5580  
GGTTTCTGCC CTGAAGGCTT CCTCCCCTCC CAATGCGGTT TAAACATAA ATAAA 5635

15 (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 base pairs  
(B) TYPE: nucleic acid  
20 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

30 ACTCTTGAGT GCCAGCGAGT AGAGTTTCTT CCTCCG 36

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
35 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

45 GCAAAGGAGC GATCCACACG AAATGTGCC 29

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
50 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

- 114 -

CTCCTCCGAG CCGCTCCGAG CTAG

24

(2) INFORMATION FOR SEQ ID NO:7:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCAAAAATGG CTGGGTGTAG GAGCAGTGTC C

31

20 (2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 base pairs

(B) TYPE: nucleic acid

25

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CGGATCCTTT ATTATAGGGG AAGTCCACGC CTAC

34

35

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

40

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

50 CGGGATCCAT CGATGAAATA TGACTACGTC CG

32

Claims

1. An adenovirus-based gene therapy vector comprising the genome of an adenovirus 2 serotype in which the Ela and Elb regions of the genome, which are involved in early stages  
5 of viral replication, have been deleted and replaced by genetic material of interest.
2. The adenovirus-based gene therapy vector of claim 1, wherein the genetic material of interest is DNA encoding cystic fibrosis transmembrane conductance regulator
- 10 3. The adenovirus-based gene therapy vector of claim 1 further comprising PGK promoter operably linked to the genetic material of interest.
4. The adenovirus-based gene therapy vector of claim 2 having substantially the same nucleotide sequence as shown in Table II (SEQ ID NO:3).
- 15 5. An adenovirus-based gene therapy vector comprising adenovirus inverted terminal repeat nucleotide sequences and the minimal nucleotide sequences necessary for efficient replication and packaging and genetic material of interest.
- 20 6. The adenovirus-based gene therapy vector of claim 5 having the adenovirus 2 sequences shown in Figure 17.
7. The adenovirus-based gene therapy vector of claim 5 further comprising PGK promoter operably linked to the genetic material of interest.
- 25 8. The adenovirus-based gene therapy vector of claim 5 in which the genetic material of interest is selected from the group consisting of DNA encoding: cystic fibrosis transmembrane conductance regulator, Factor VIII, and Factor IX.
- 30 9. An adenovirus-based gene therapy vector comprising an adenovirus genome which has been deleted for all E4 open reading frames, except open reading frame 6, and additionally comprising genetic material of interest.
10. The adenovirus-based gene therapy vector of claim 9 further comprising PGK  
35 promoter operably linked to the genetic material of interest.
11. The adenovirus-based gene therapy vector of claim 9 in which the Ela and Elb regions of the genome, which are involved in early stages of viral replication, have been deleted.

12. The adenovirus-based gene therapy vector of claim 9 in which the E3 region has been deleted.
13. An adenovirus-based gene therapy vector comprising an adenovirus genome which  
5 has been deleted for all E4 open reading frames, except open reading frame 3, and additionally comprising genetic material of interest.
14. The adenovirus-based gene therapy vector of claim 13 in which the Ela and Elb  
10 regions of the genome, which are involved in early stages of viral replication, have been deleted.
15. The adenovirus-based gene therapy vector of claim 13 further comprising PGK promoter operably linked to the genetic material of interest.
- 15 16. The adenovirus-based gene therapy vector of claim 13 in which the E3 region has been deleted.
17. A method for treating or preventing cystic fibrosis in a patient comprising  
20 administering to the pulmonary airways of the patient, a gene therapy vector comprising DNA encoding cystic fibrosis transmembrane conductance regulator.
18. The method of claim 17 wherein the gene therapy vector is an adenovirus-based gene  
therapy vector comprising the genome of an adenovirus 2 serotype in which the Ela and Elb  
25 regions of the genome, which are involved in early stages of viral replication, have been deleted and replaced by DNA encoding cystic fibrosis transmembrane conductance regulator.
19. The method of claim 17 wherein the gene therapy vector further comprises PGK  
30 promoter operably linked to the DNA encoding cystic fibrosis transmembrane conductance regulator.
20. The method of claim 17 wherein the gene therapy vector is an adenovirus-based gene  
therapy vector comprising adenovirus inverted terminal repeats and the minimal sequences  
necessary for efficient replication and packaging and DNA encoding cystic fibrosis  
35 transmembrane conductance regulator.
21. The method of claim 20 wherein the gene therapy vector further comprises PGK promoter operably linked to the DNA encoding cystic fibrosis transmembrane conductance regulator.

22. The method of claim 17 wherein the gene therapy vector is an adenovirus-based gene therapy vector comprising an adenovirus genome which has been deleted for all E4 open reading frames, except open reading frame 6, and additionally comprising DNA encoding  
5 cystic fibrosis transmembrane conductance regulator.

23. The method of claim 22 wherein the gene therapy vector further comprises PGK promoter operably linked to the DNA encoding cystic fibrosis transmembrane conductance regulator.  
10

24. The method of claim 17 wherein the gene therapy vector is an adenovirus-based gene therapy vector comprising an adenovirus genome which has been deleted for all E4 open reading frames, except open reading frame 6, and has been deleted for the Ela and Elb regions of the genome, which are involved in early stages of viral replication, and additionally  
15 comprising DNA encoding cystic fibrosis transmembrane conductance regulator.

25. The method of claim 24 wherein the gene therapy vector further comprises PGK promoter operably linked to the DNA encoding cystic fibrosis transmembrane conductance regulator.



1/50

## PARTIAL cDNA CLONES OF THE CFTR GENE

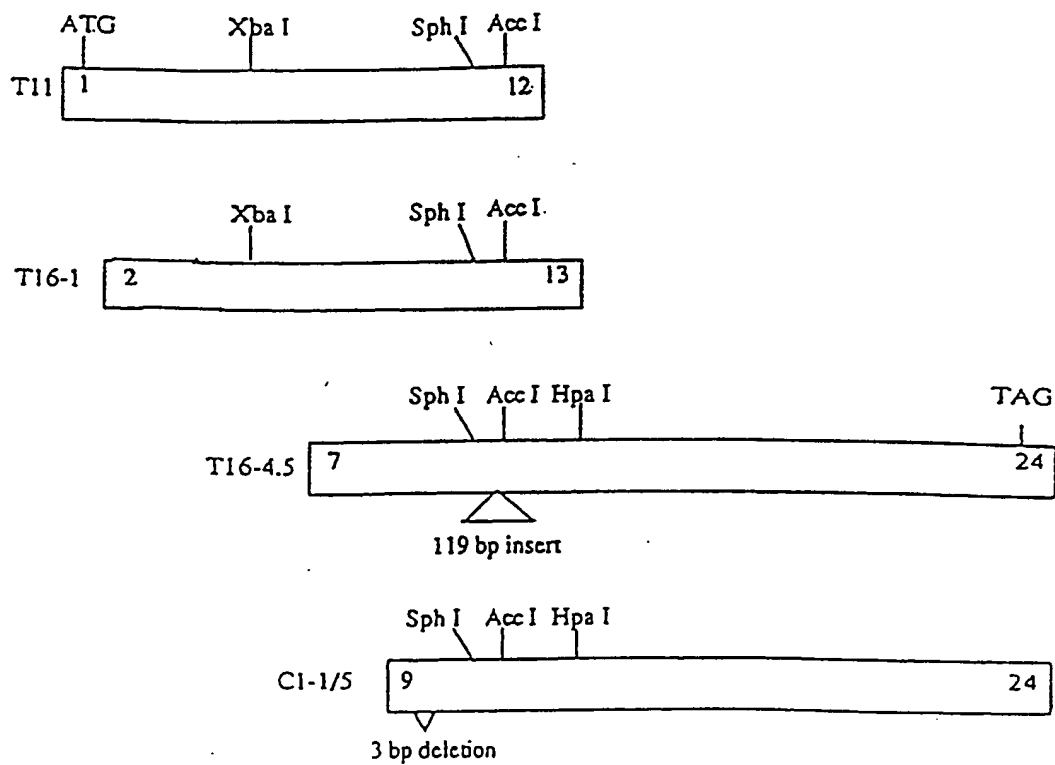


Figure 1

2/50

## STRATEGY FOR CONSTRUCTING pKK- CFTR I

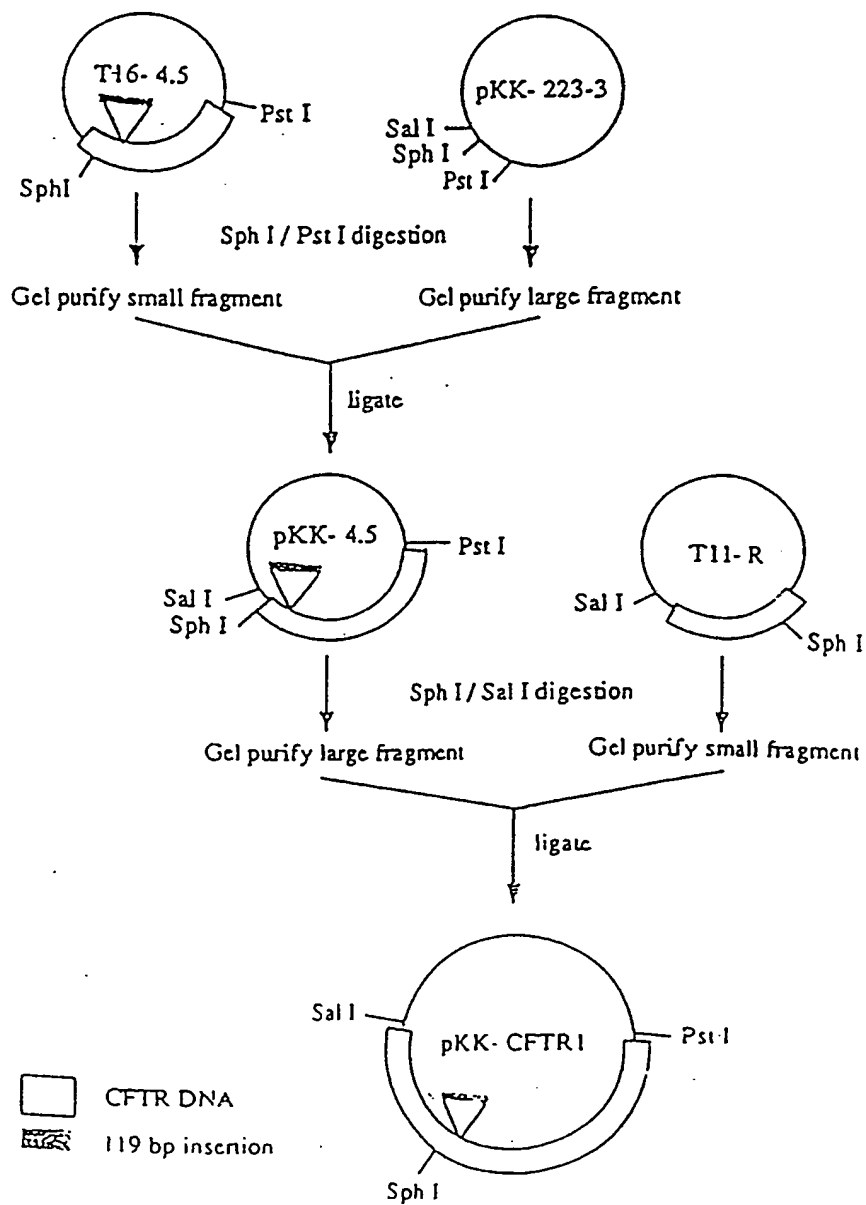


Figure 2

SUBSTITUTE SHEET (RULE 26)

3/50

## CONSTRUCTION OF THE pKK- CFTR2 PLASMID

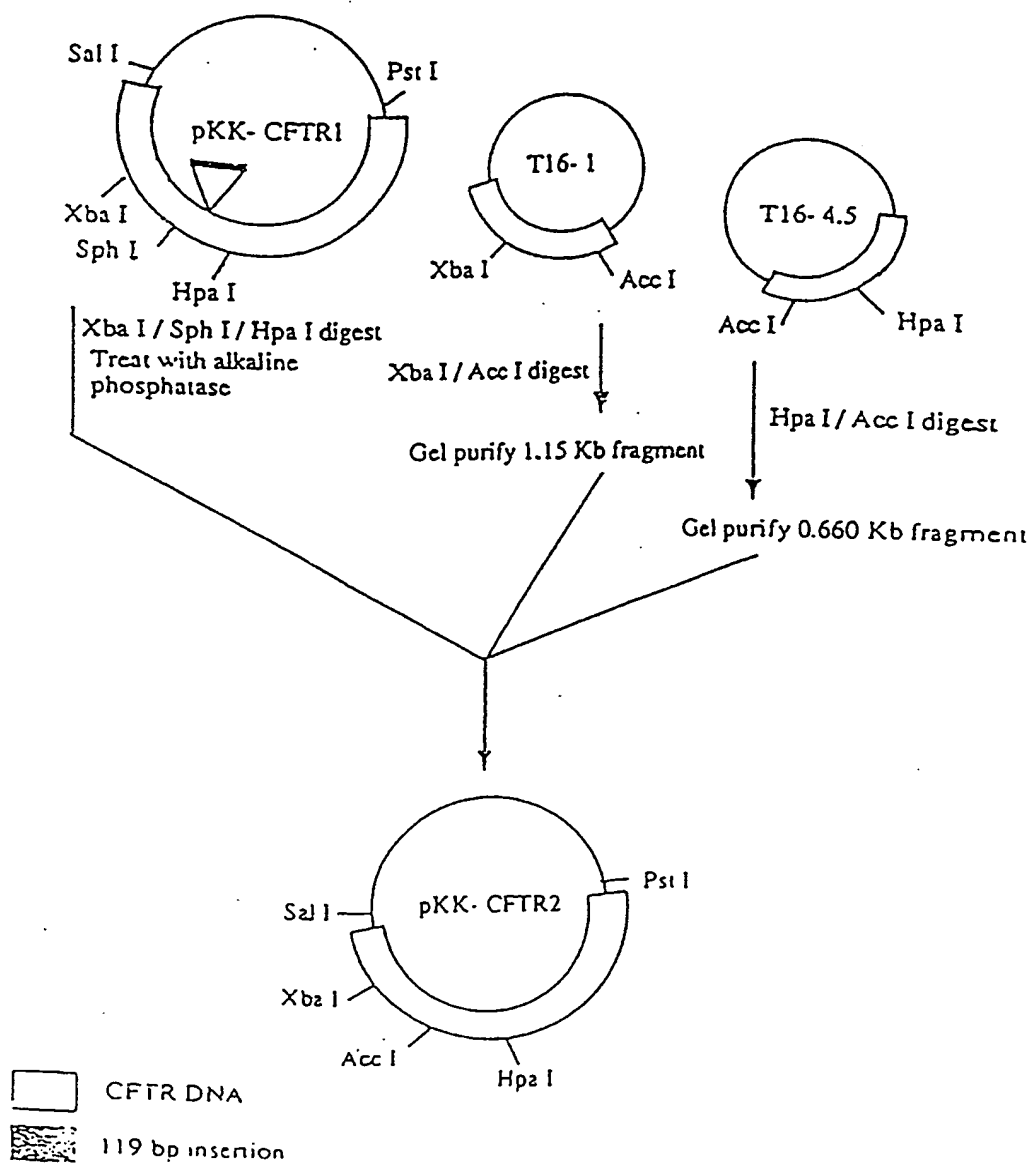


Figure 3

## STRATEGY FOR CONSTRUCTING THE pSC- CFTR2 PLASMID

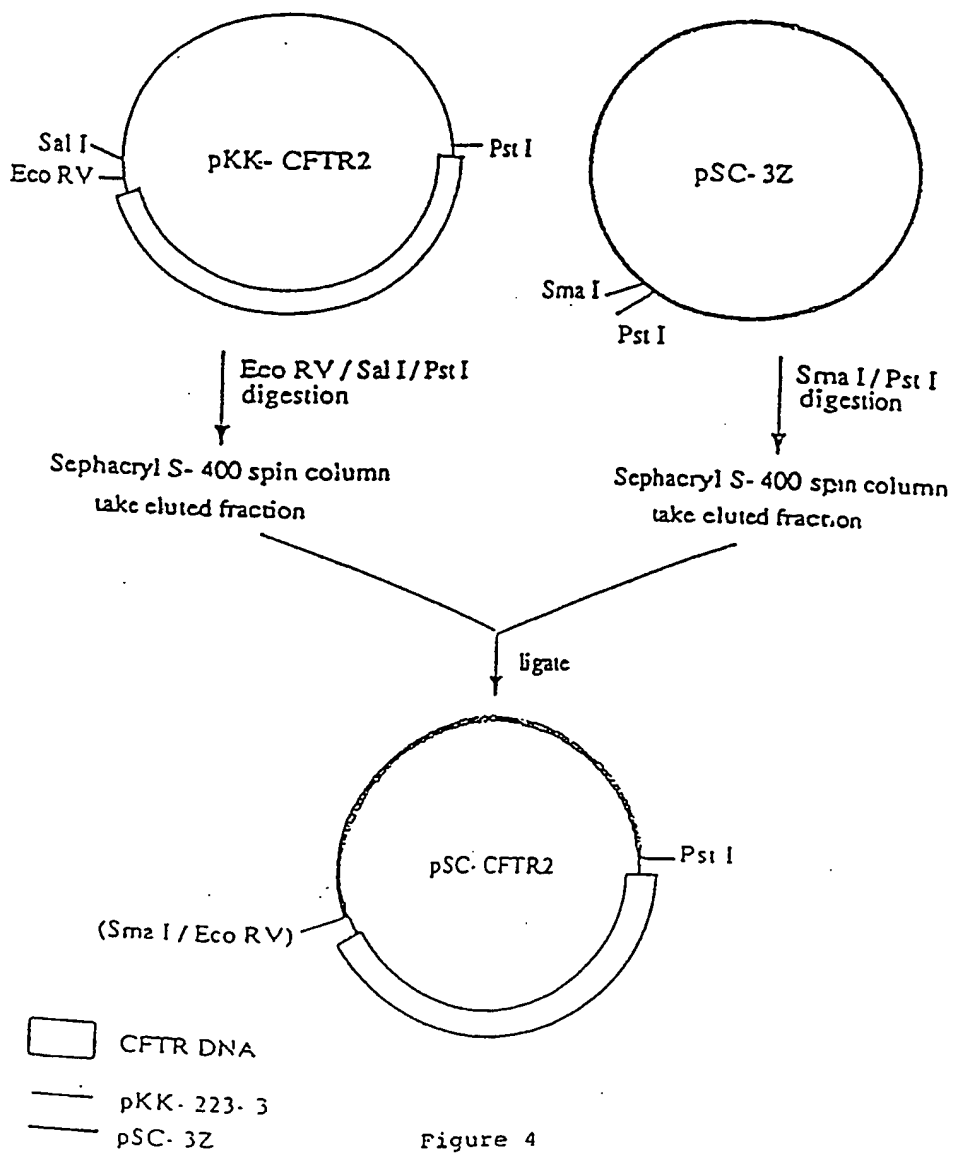


Figure 4

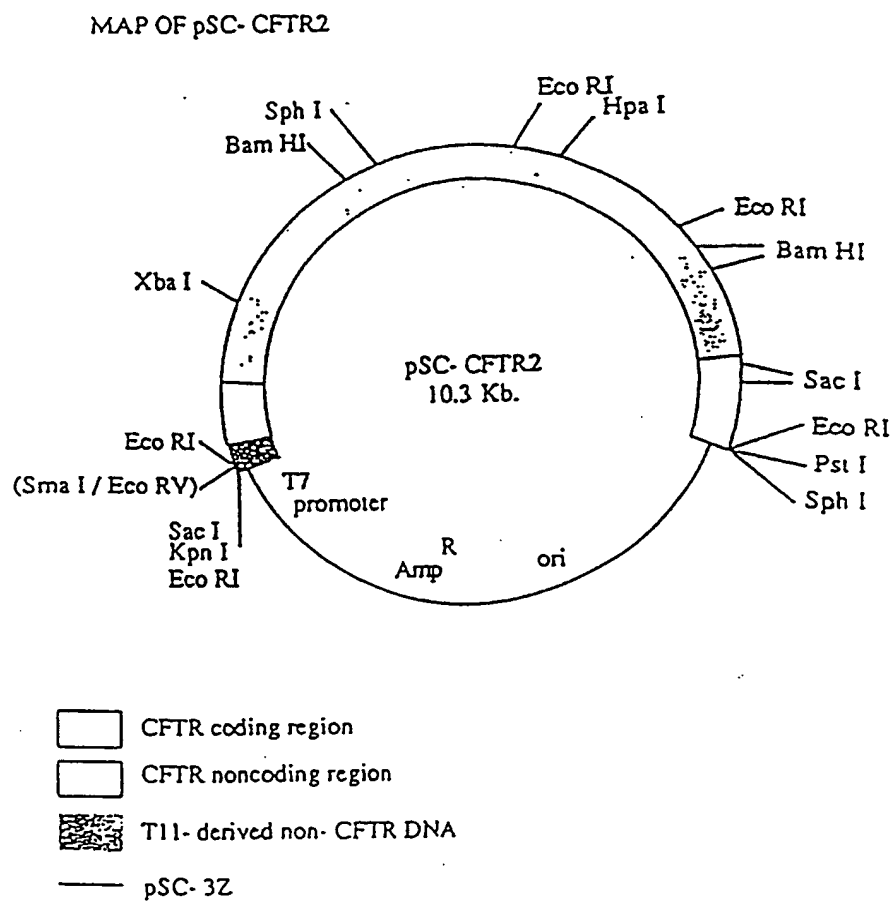


Figure 5

6/50

S                    bp 1716  
 P                    |  
 h                    |-----x-----xSynthetic Intron-----x-----x  
 i                    |  
 |-----1195RG-----  
 CCAACTAGAAAGAGGTAAGGGGCTCACCAGTTCAAAATCTGAAGTGGAGACAGGAC  
 GTACGGTTGATCTTCTCCATTCCCCGAGTGGTCAAGTTTTAGACTTCACCTCTGTCCTG  
 <-----1198RG-----  
  
 bp 1717  
 =====|  
 |  
 ----->|-----  
 CTGAGGTGACAATGACATCTACTCTGACATTCTCTCCTCAGGACATCTCCAAGTTTGCAG  
 GACTCCACTGTTACTGTAGATGAGACTGTAAGAGAGGAGTCCTGTAGAGGTTCAAACGTC  
 -----|<-----1197RG-----  
  
 H  
 i  
 n  
 c  
 I  
 I  
  
 -----1196RG----->  
 AGAAAGACAATATAGTTCTTGGAGAAAGGTGGAATCACACTGAGTGGAGGTC  
 TCTTCTGTATATCAAGAACCTCTTCCACCTTAGTGTGACTCACCTCCAG  
 -----|

Figure 6

7/50

## CONSTRUCTION OF THE pKK- CFTR3 cDNA

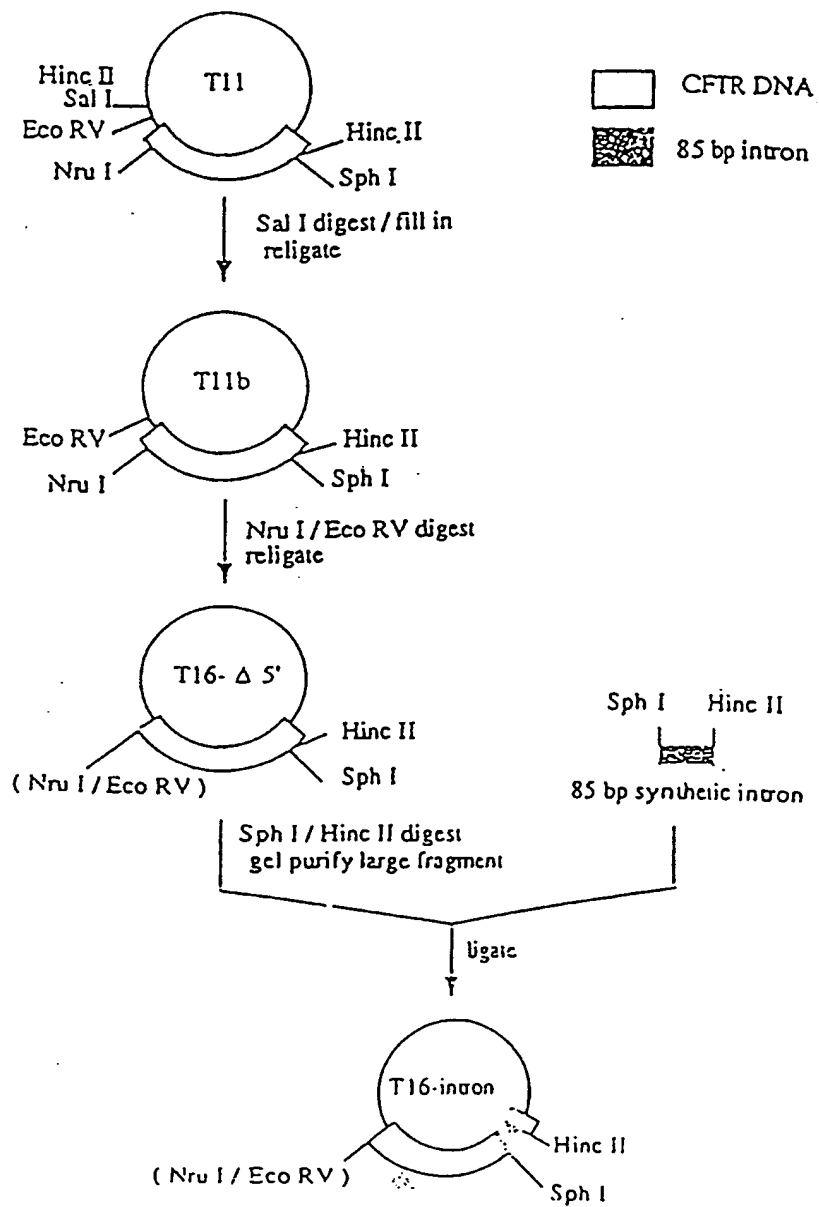


Figure 7A

8/50

## CONSTRUCTION OF THE pKK-CFTR3 CLONE (cont'd.)

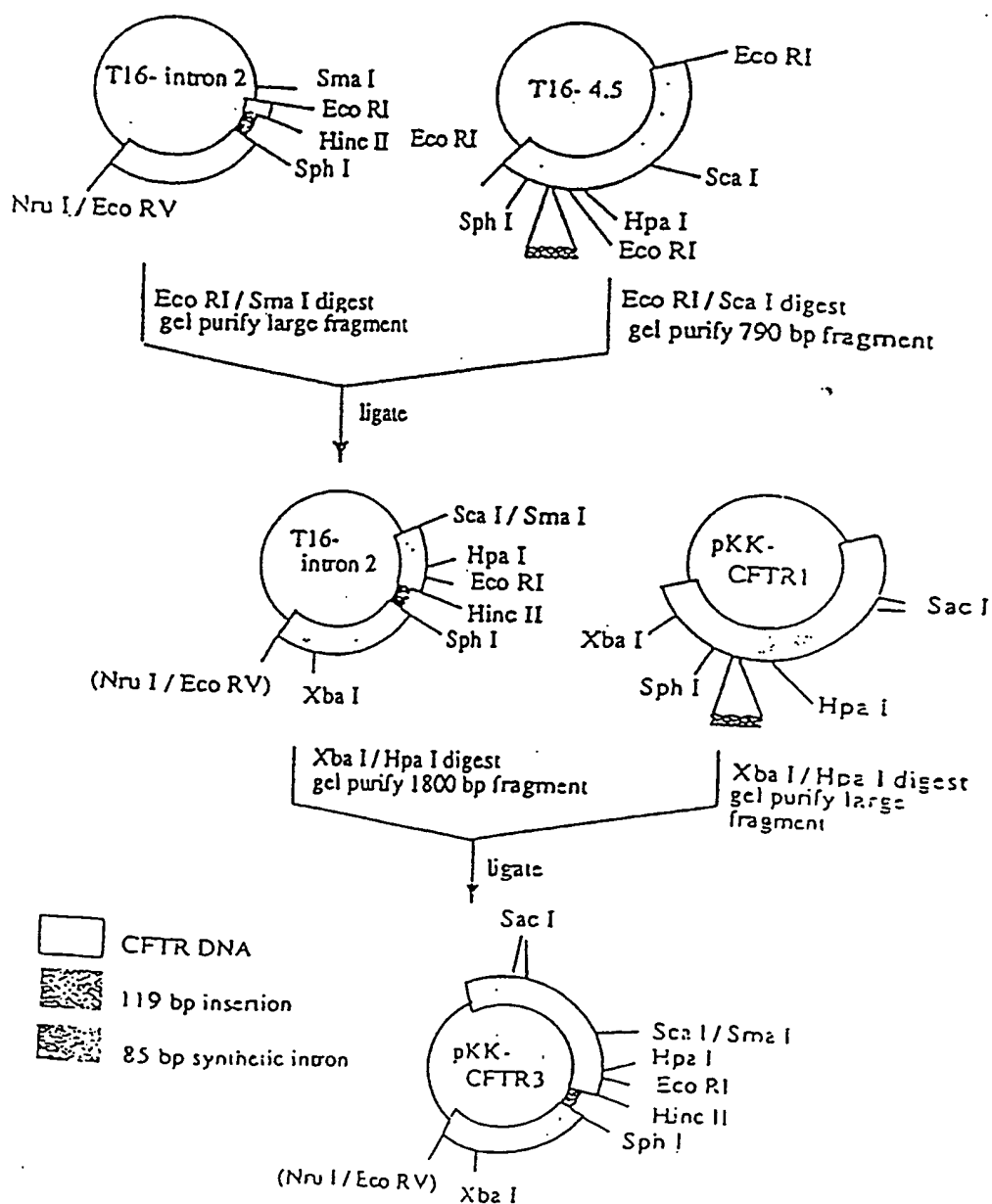


Figure 7B



## MAP OF pKK- CFTR3

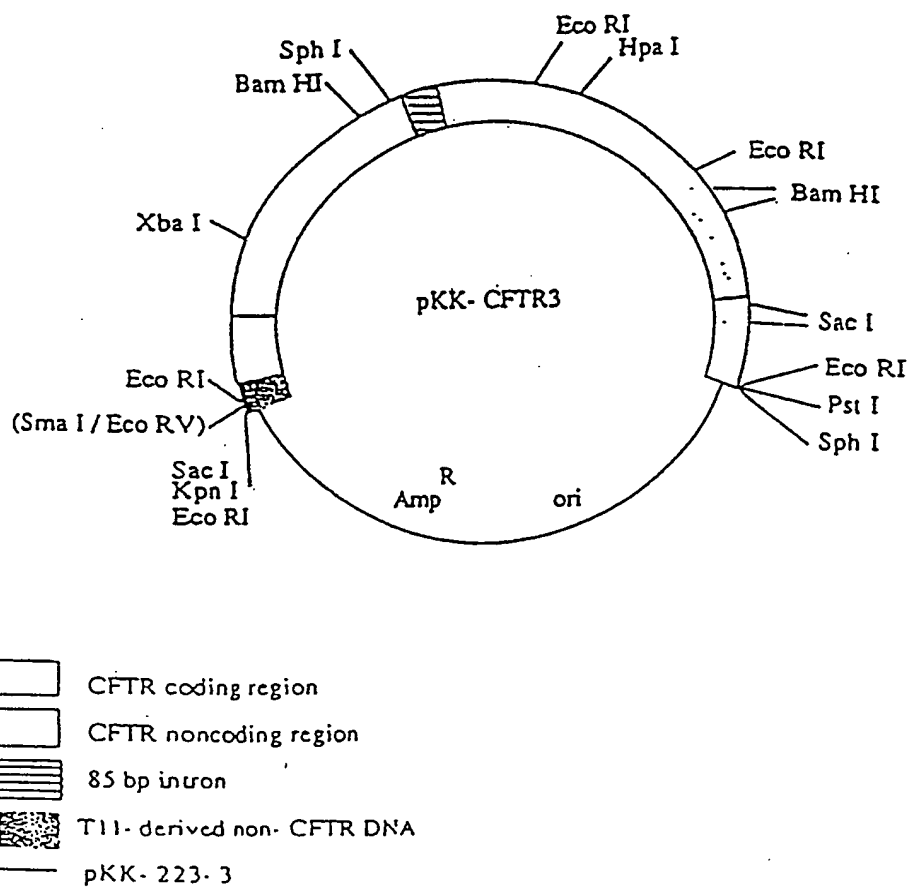


Figure 8



11/50



Figure 10A

Figure 10B

12/50

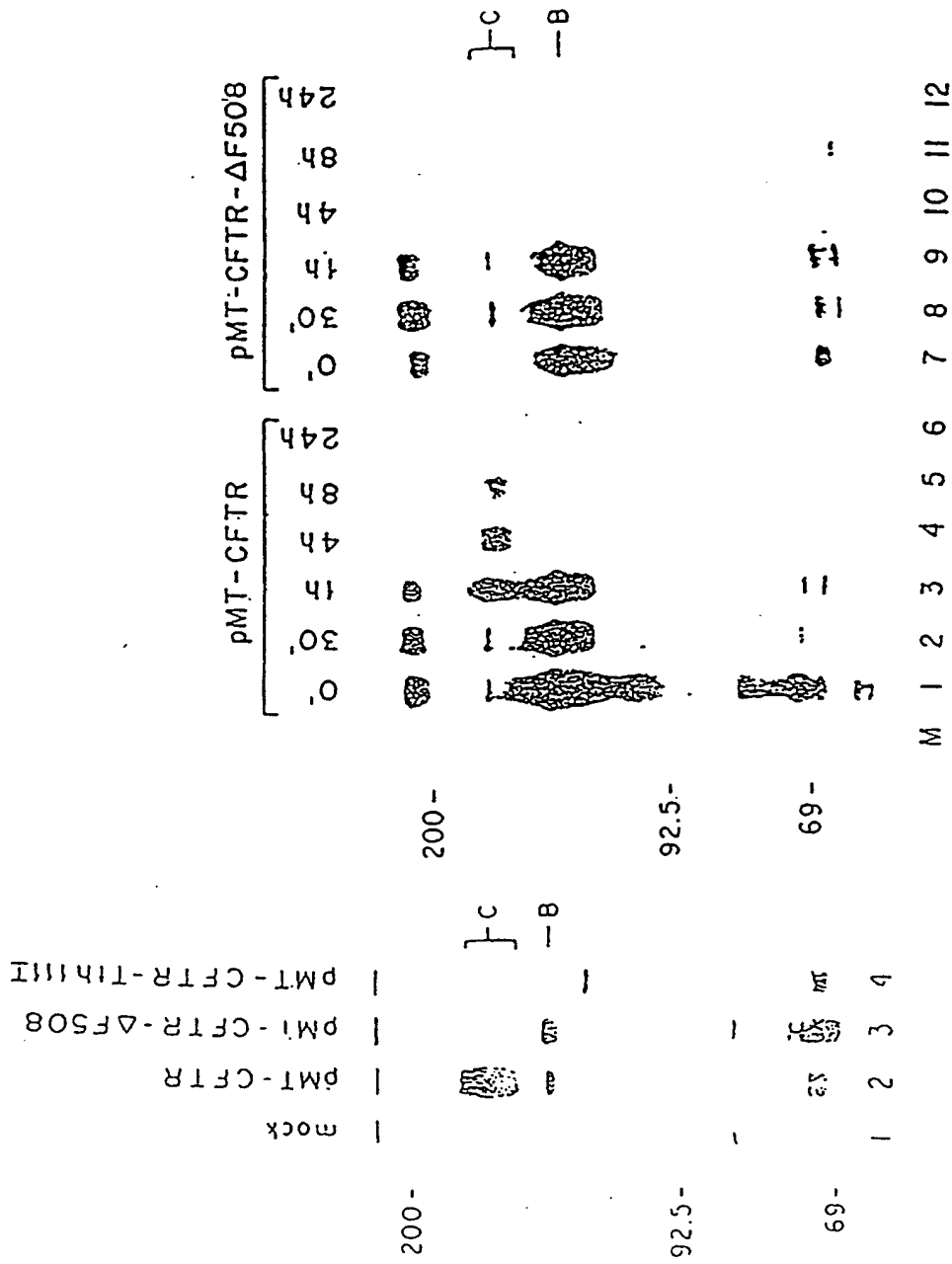


Figure 11A

Figure 11B

Figure 12A

Figure 12B

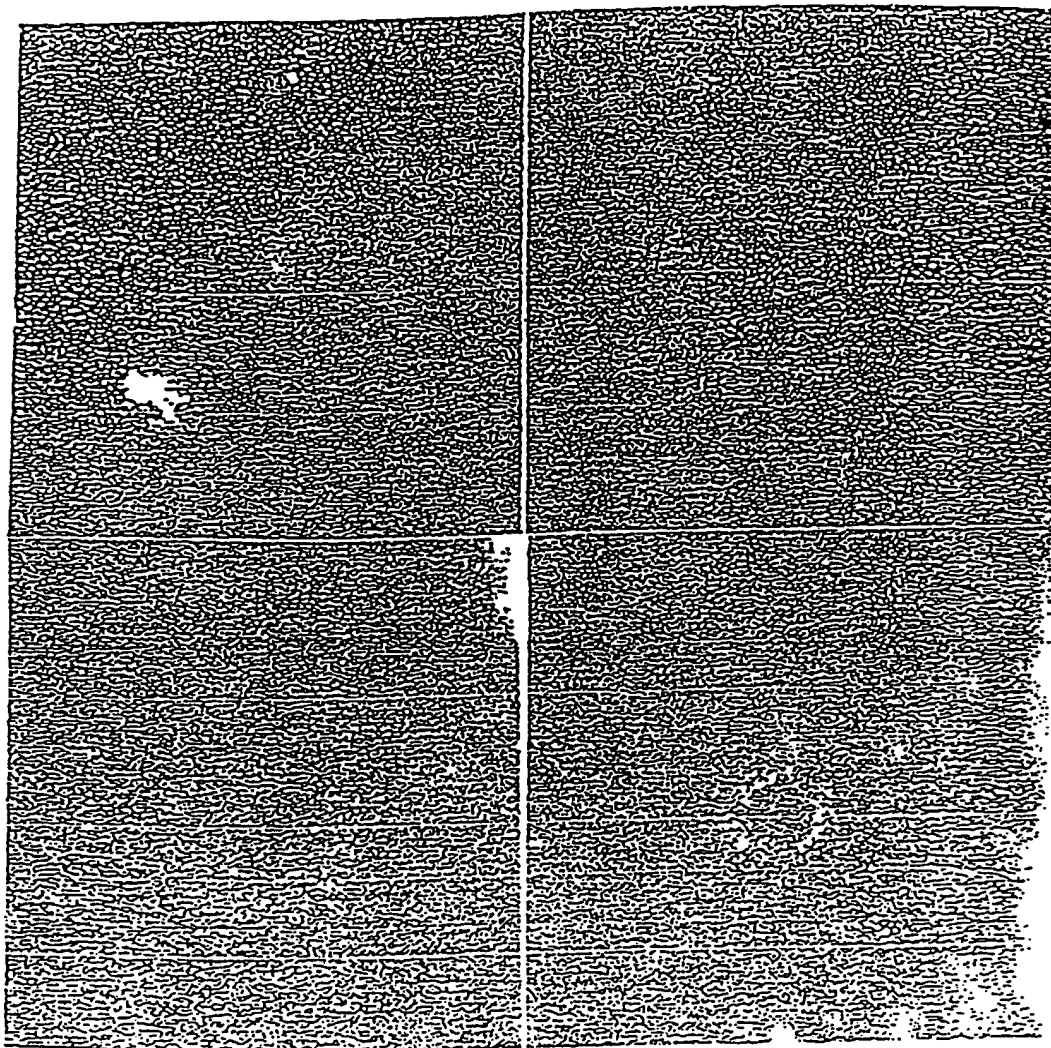


Figure 12C

Figure 12D

14/50

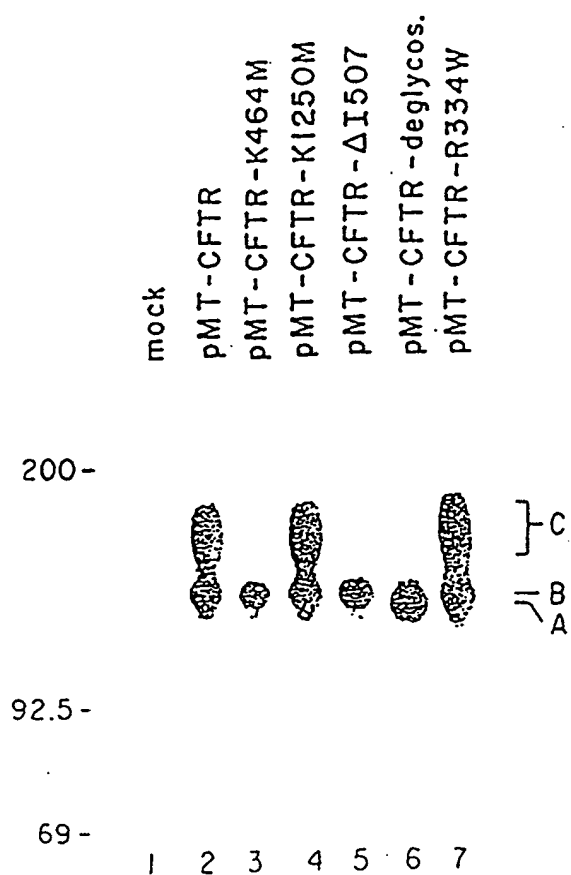
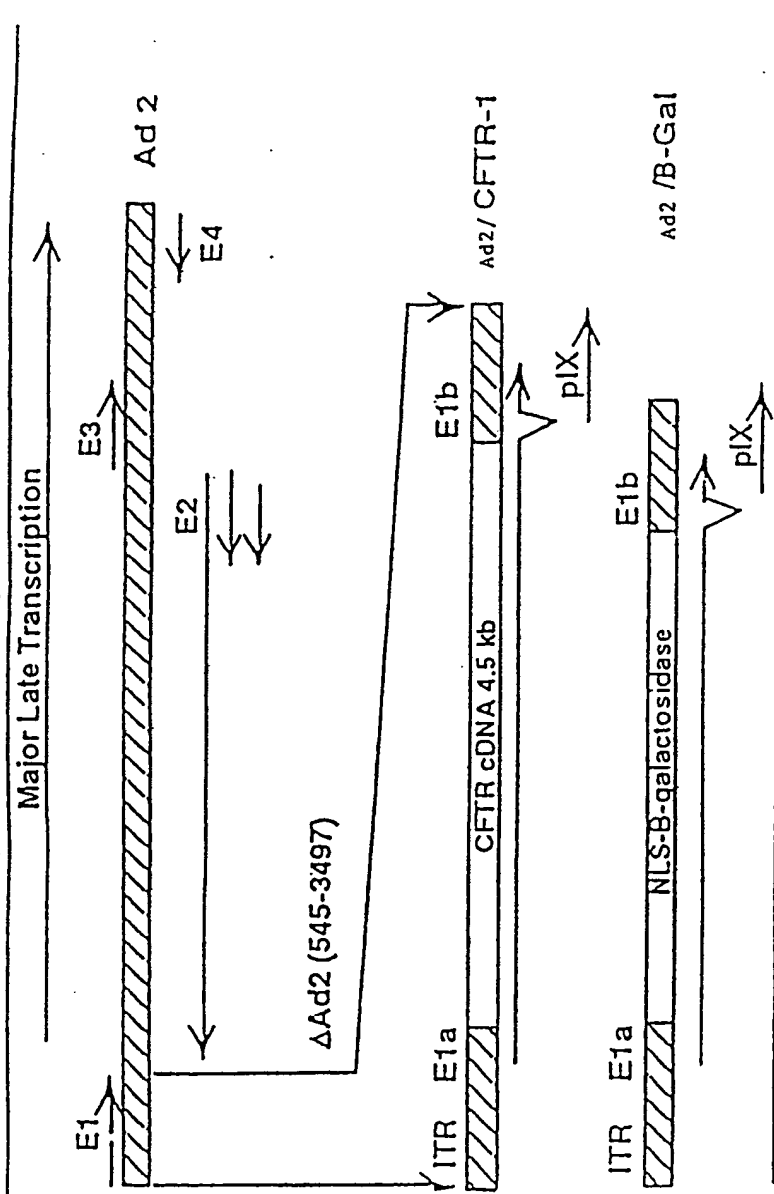


Figure 13

### FIGURE 1 MAP OF VECTOR



**Figure 14**

16/50

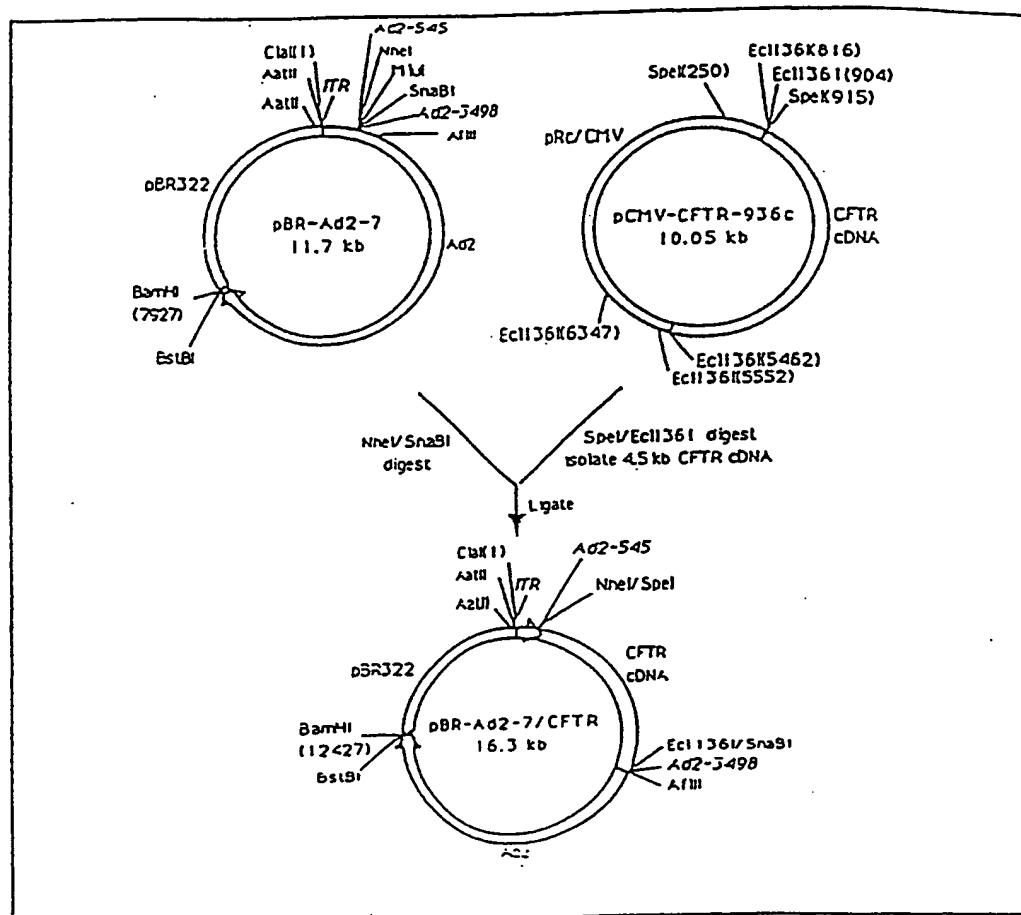


Figure 15



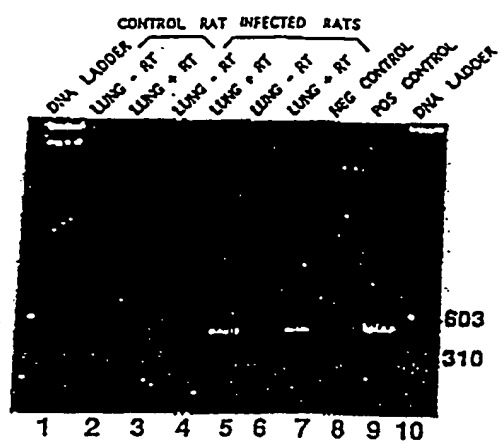


Figure 16

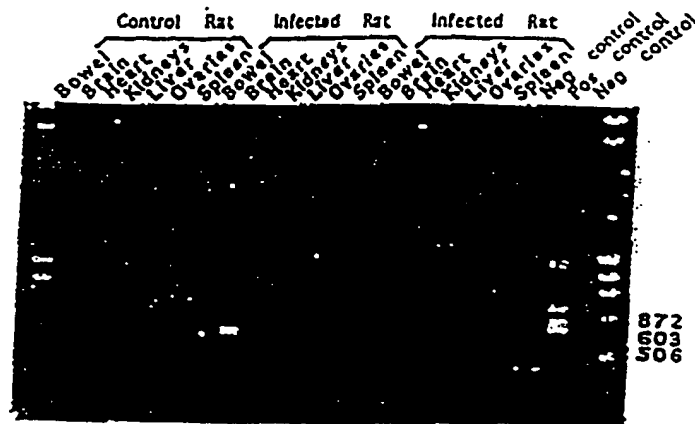


Figure 17

19/50

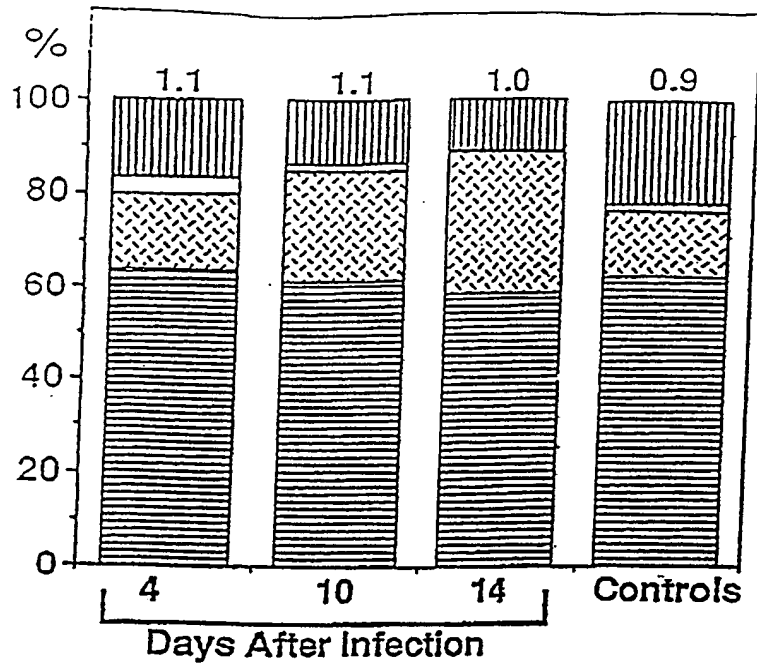


Figure 18A

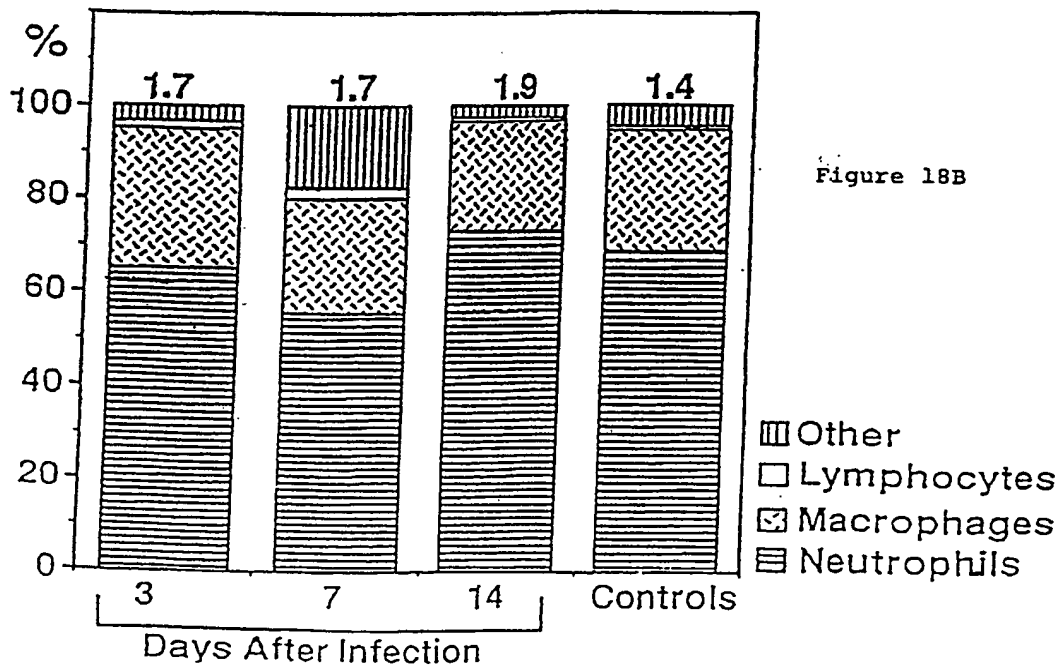


Figure 18B



Figure 19

21/50

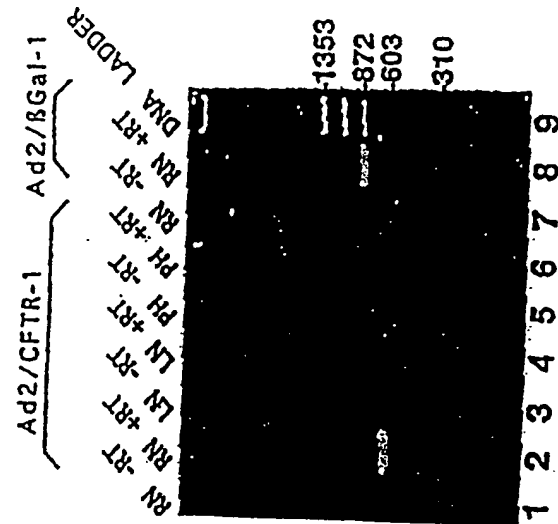


Figure 20A

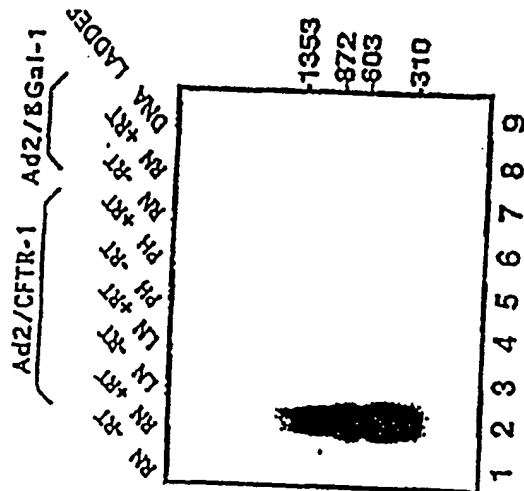


Figure 20B

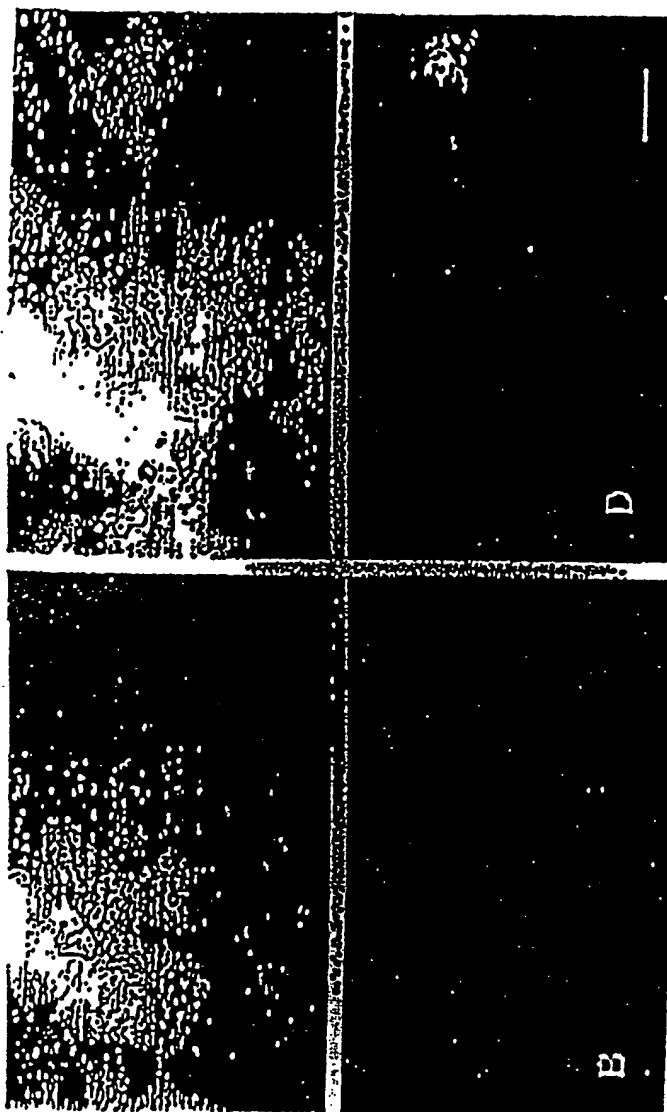


Figure 21

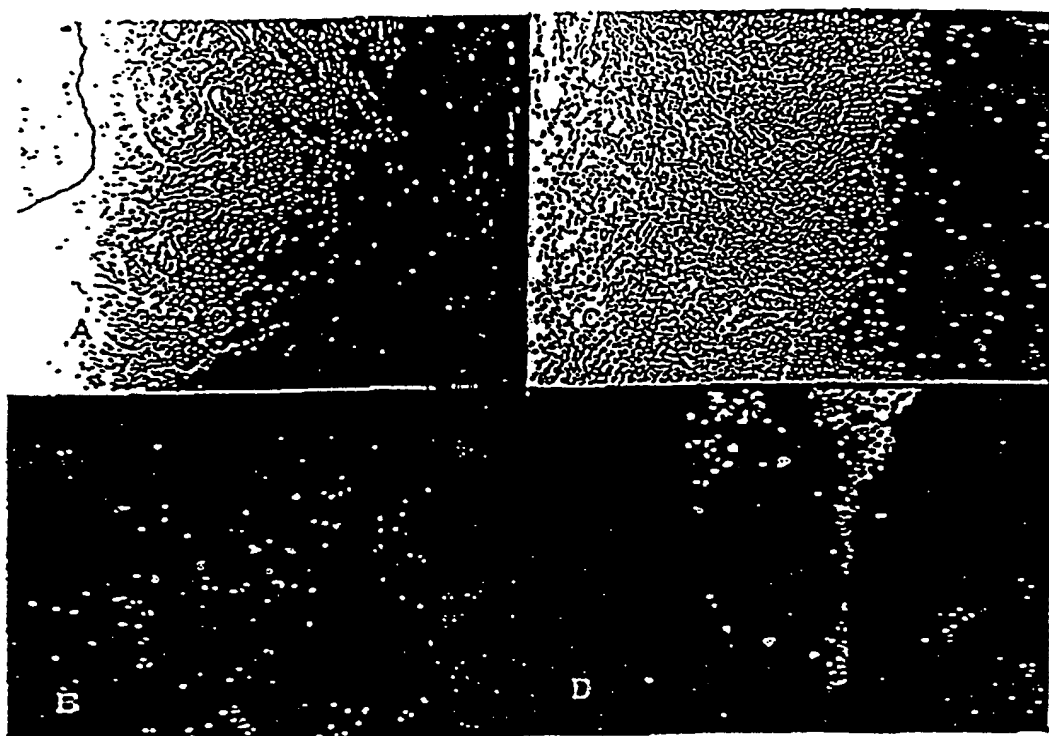
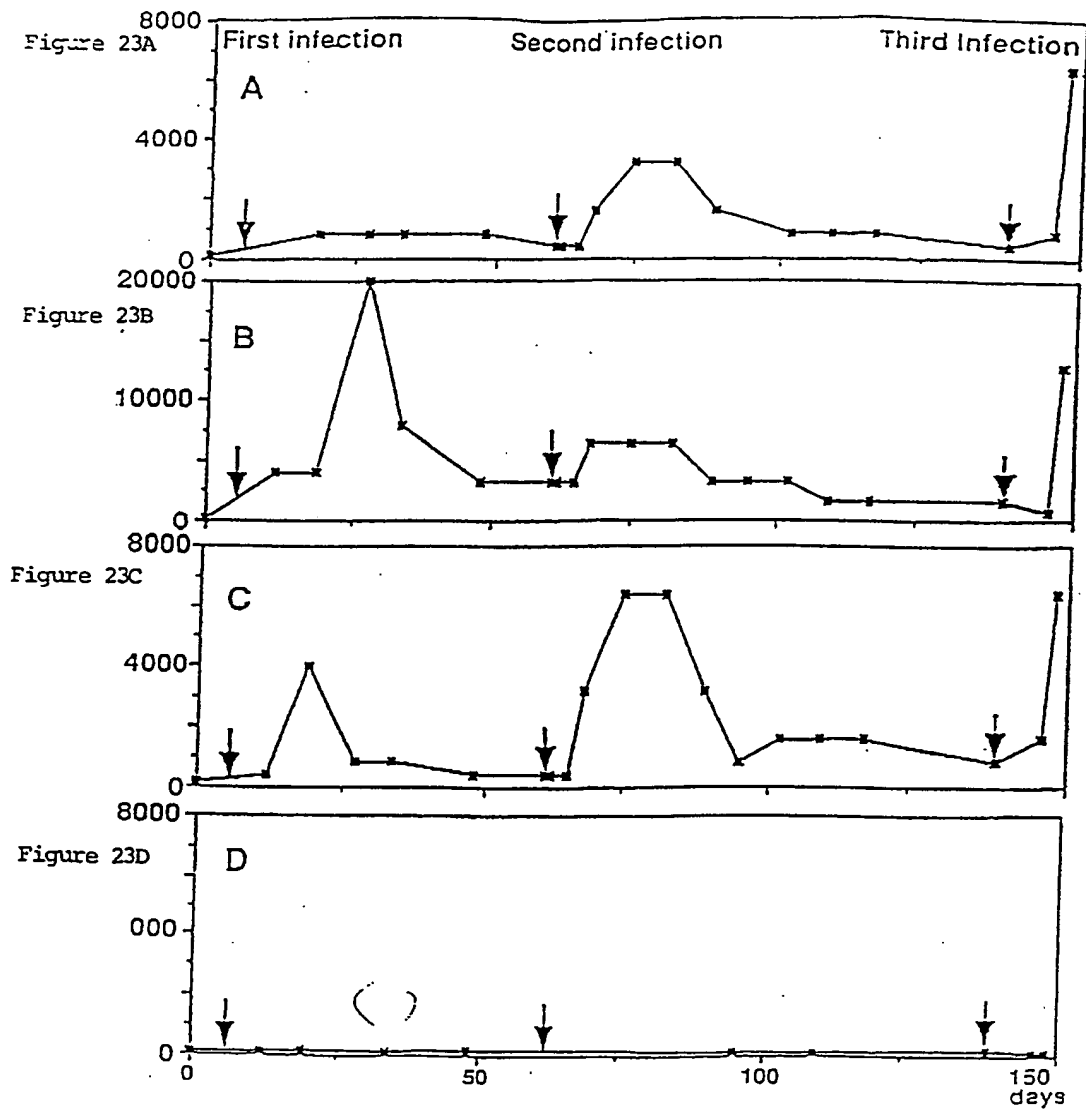


Figure 22

24/50

## ANTIBODY TITERS





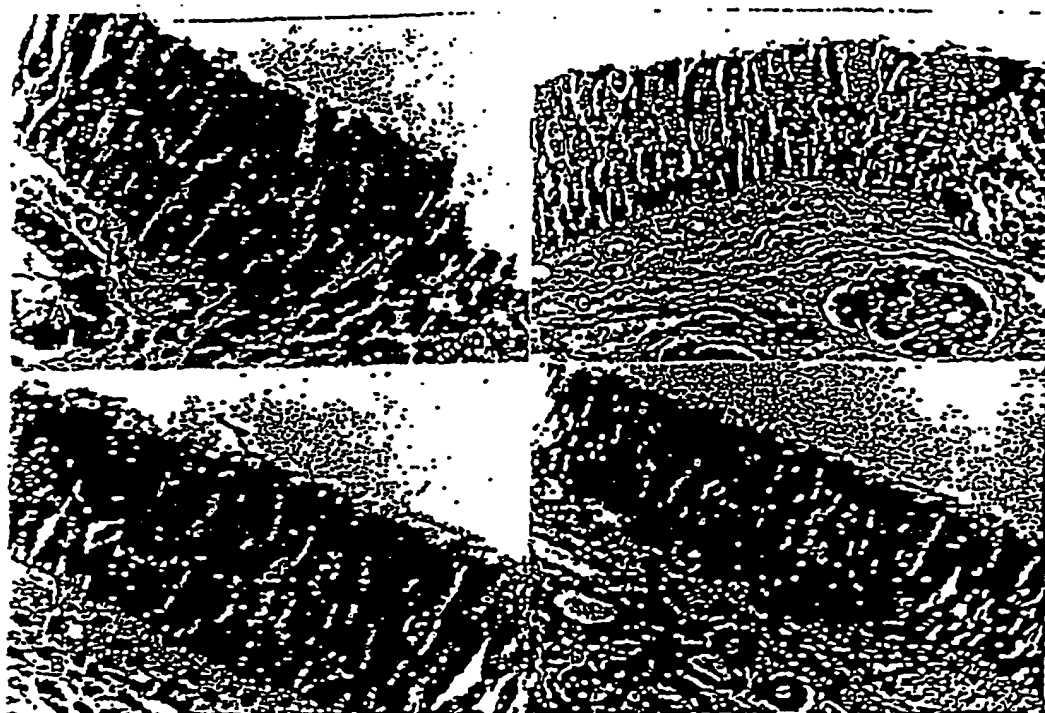


Figure 24

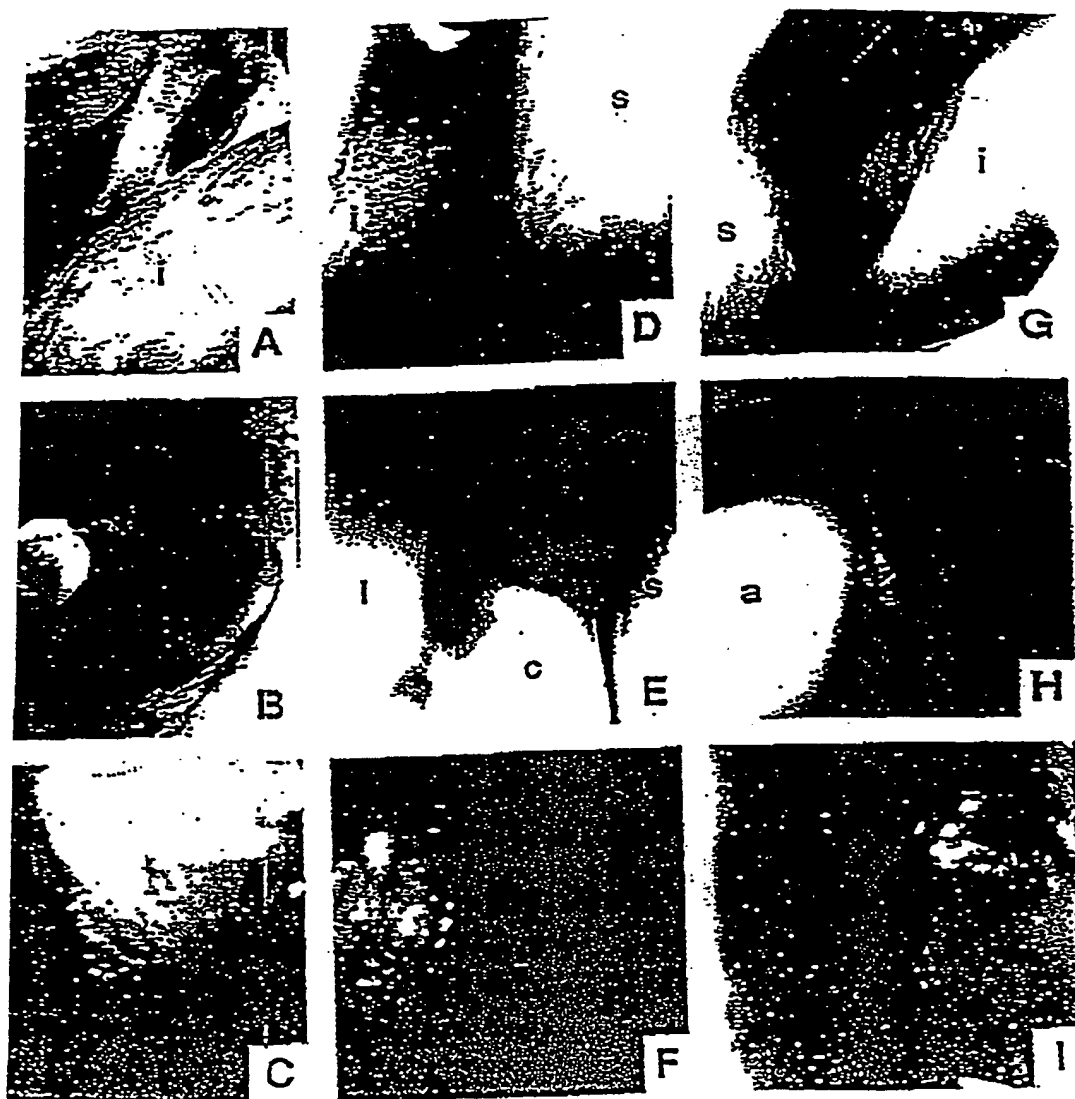


Figure 25



Figure 26

28/50

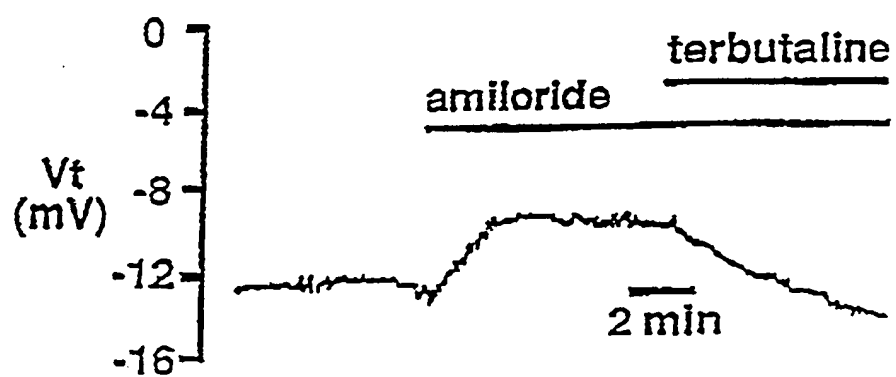


Figure 27

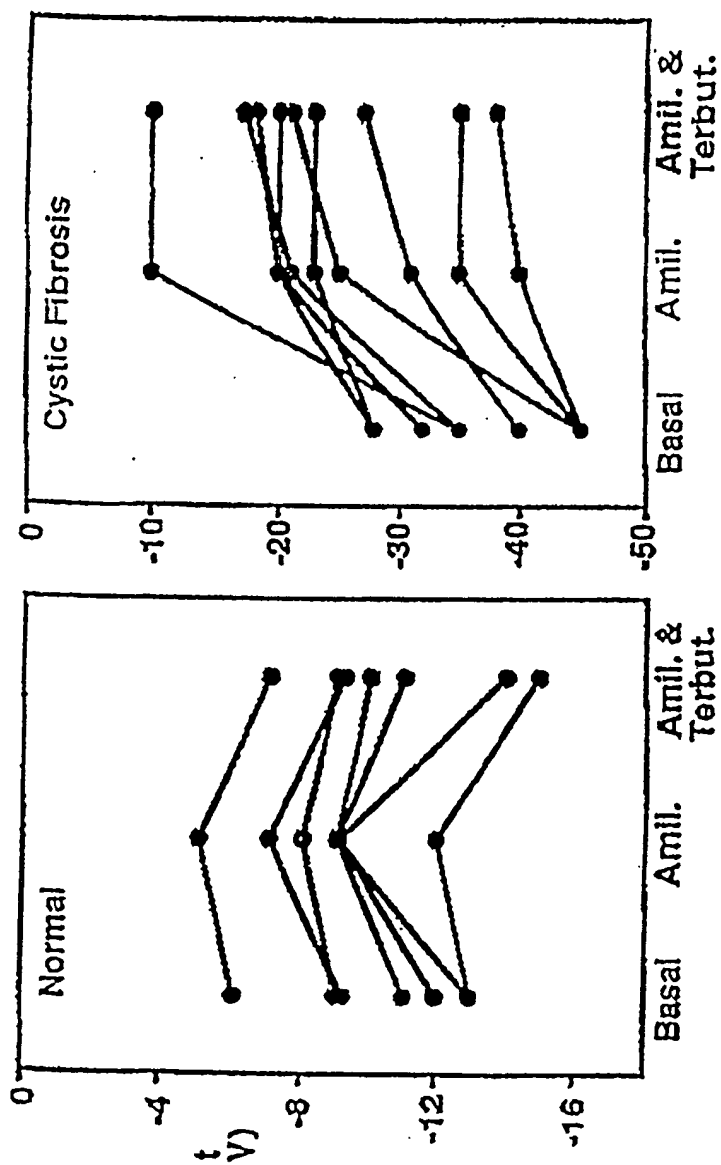
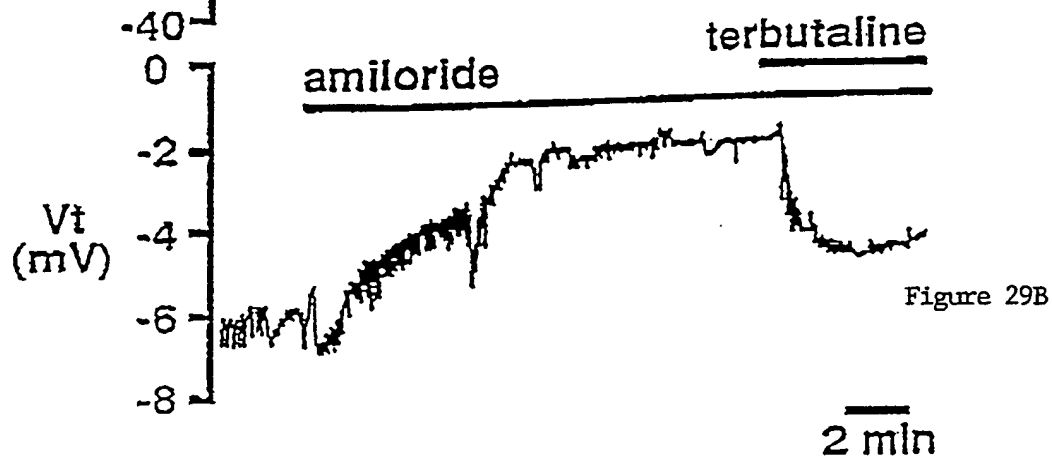
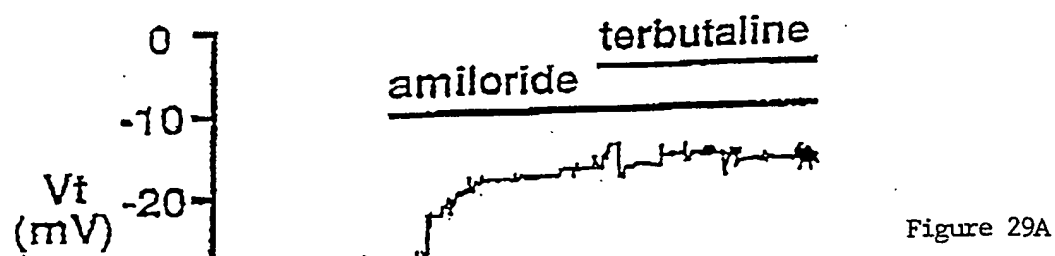


Figure 28B

Figure 28A

30/50





32/50

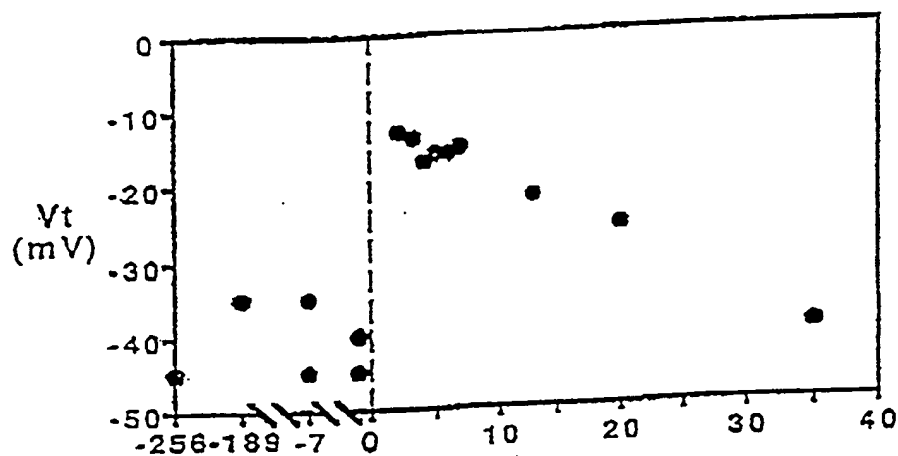


Figure 30C

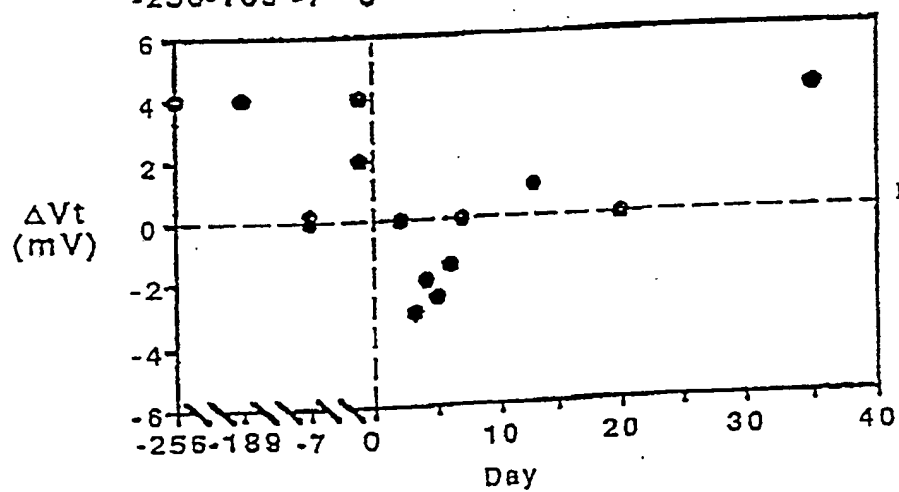
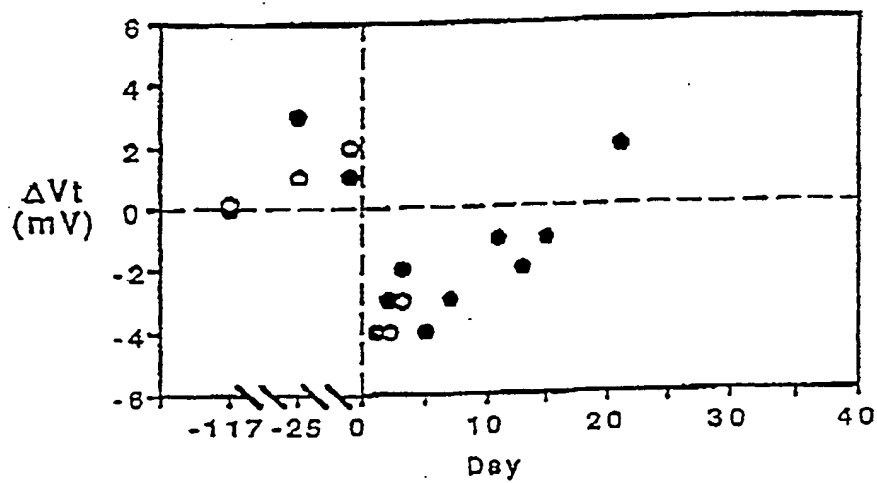
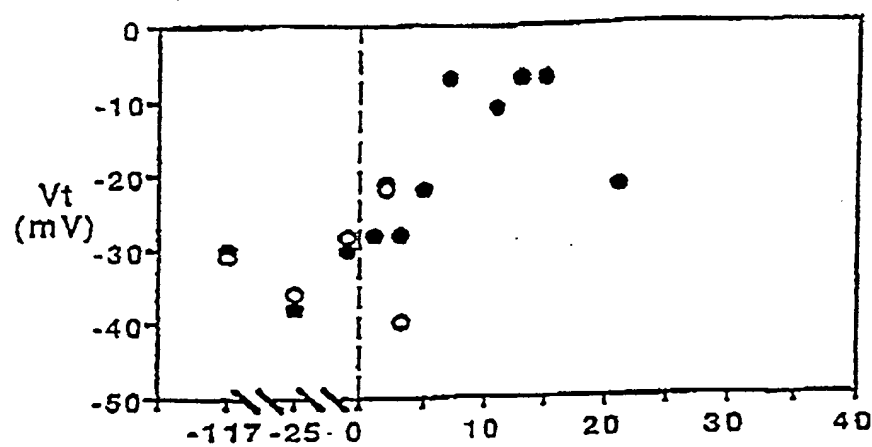


Figure 30D





34/50

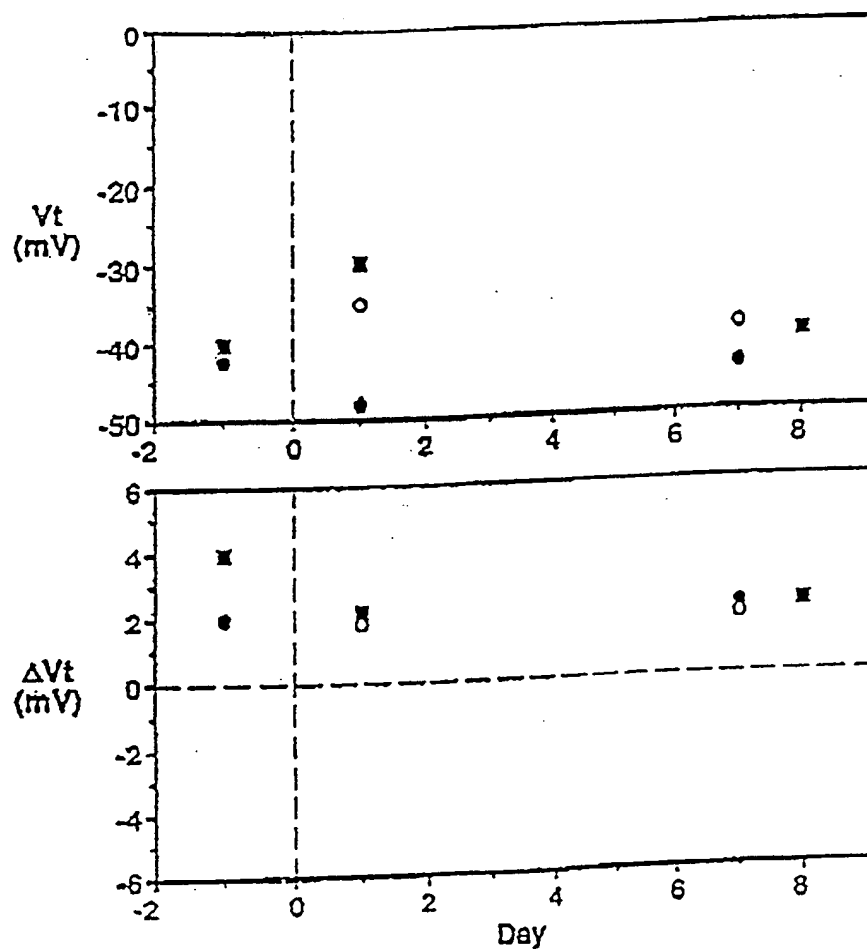
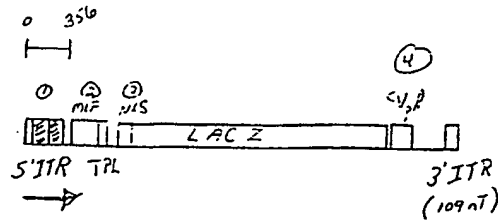


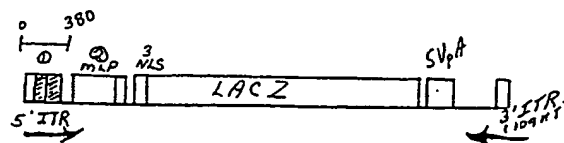
Figure 31

35/50



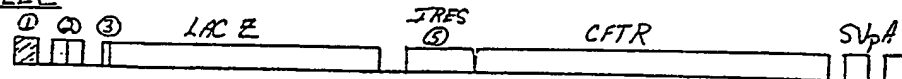
- ① Adenovirus Type 2 packaging signal and E1 enhancer region  
 ② Adenovirus Type 2 major Late Promoter and Tri-partite leader  
 ③ SV40 Transgene nuclear Localization Signal  
 ④ SV40 Poly Adenylation Signal

### PAVII

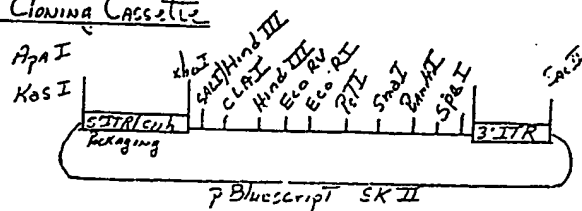


- ① Adenovirus Type 2 packaging signal and E1 enhancer region  
 ② Adenovirus Type 2 major Late Promoter and Tri-partite leader  
 ③ SV40 Transgene nuclear Localization Signal  
 ④ SV40 Poly Adenylation Signal

### PAV I/II LEC



- ⑤ EMC VIRUS Internal Ribosomal entry site - for polycistronic Translation  
 PAVI Cloning Cassette



### Expression Cassette

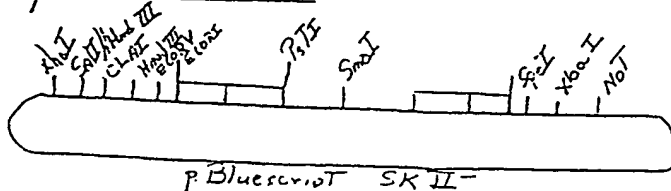


Figure 32

SUBSTITUTE SHEET (RULE 26)

36/50

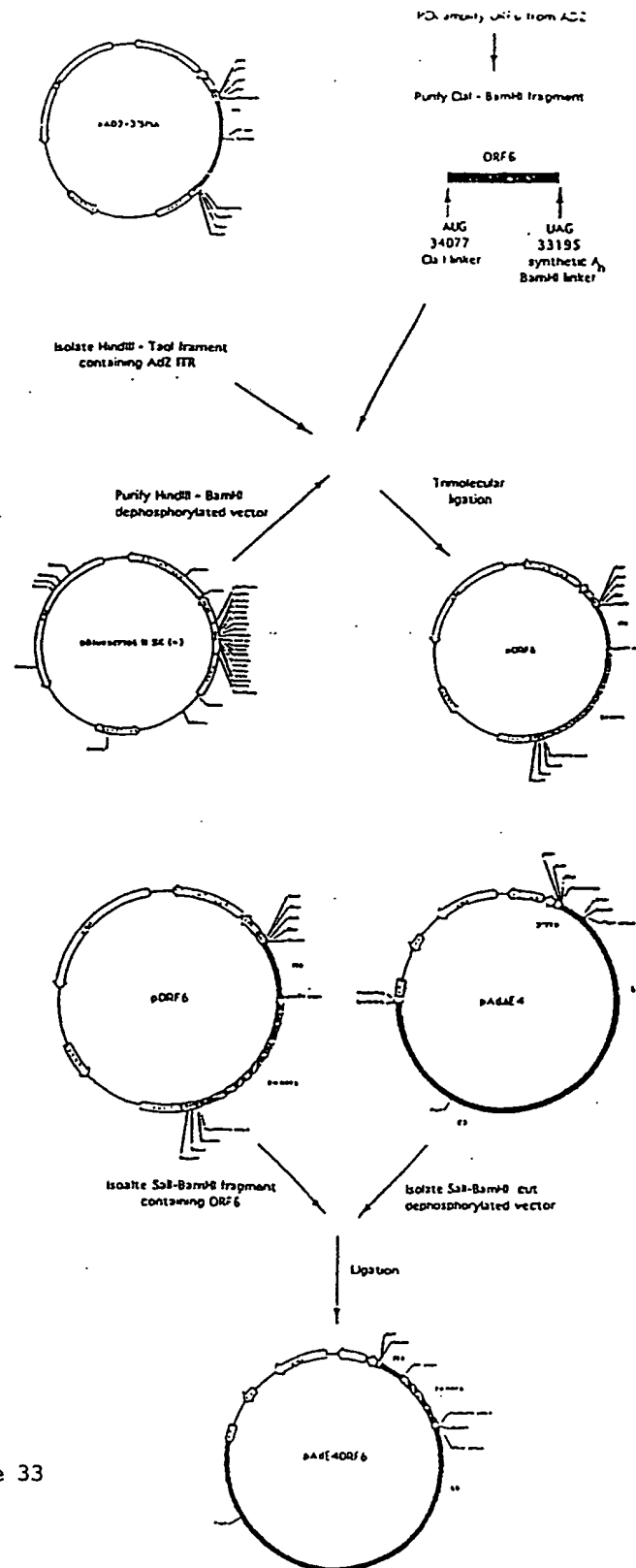


Figure 33

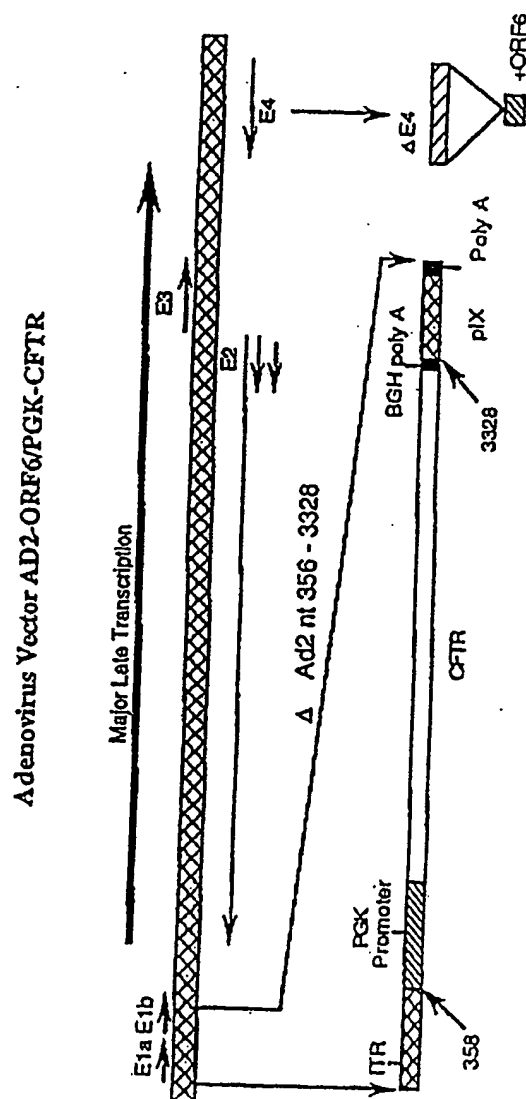


Figure 34

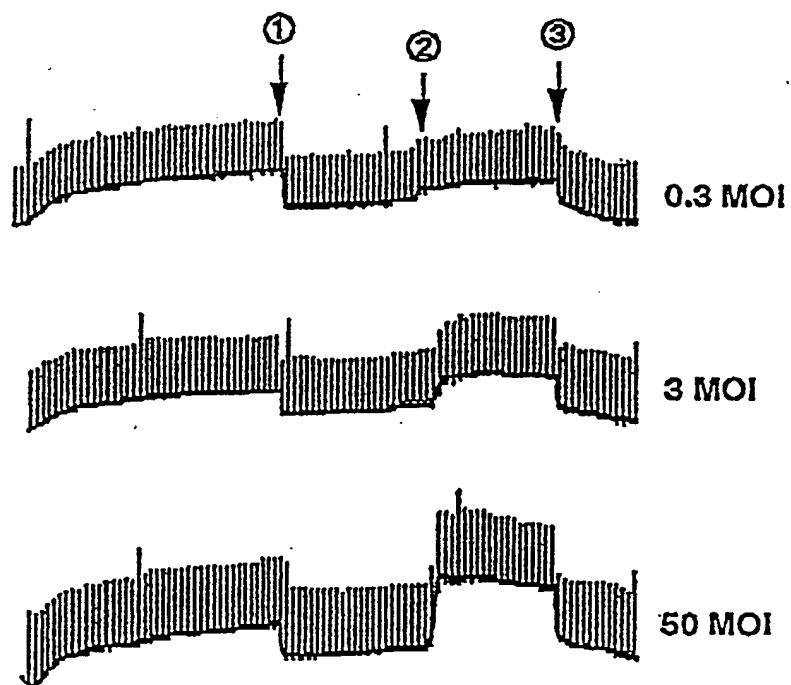


Figure 35

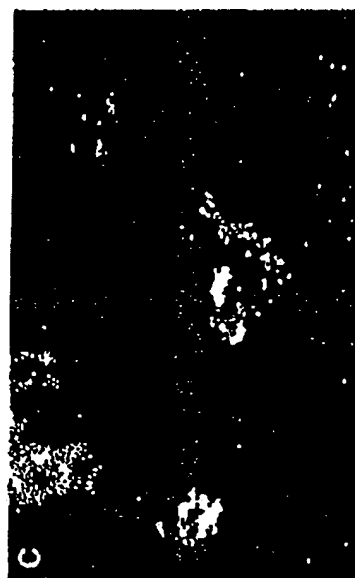


Figure 36 C

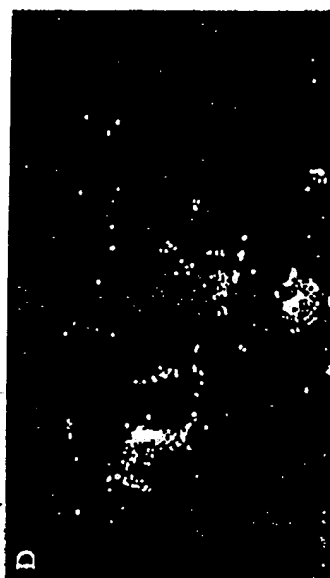


Figure 36D

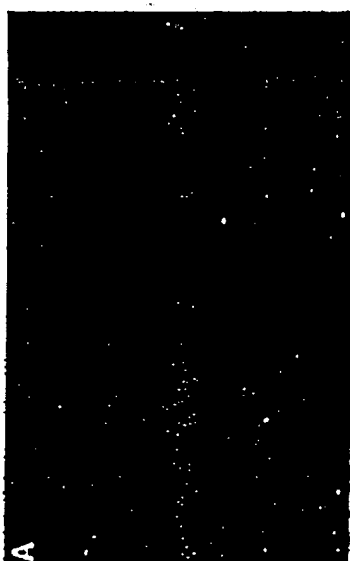


Figure 36A

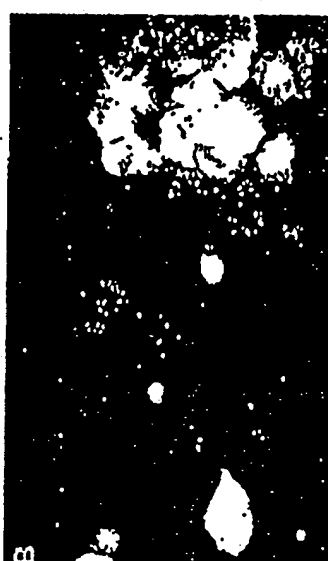


Figure 36B

40/50

Figure 37C

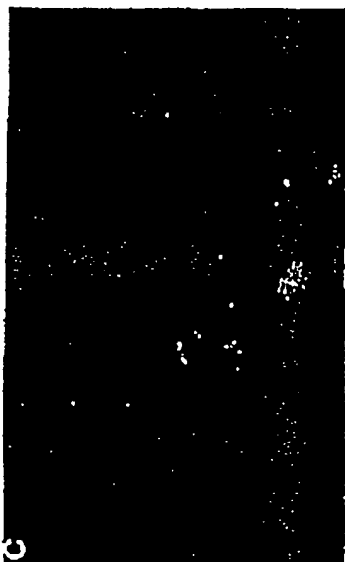


Figure 37D

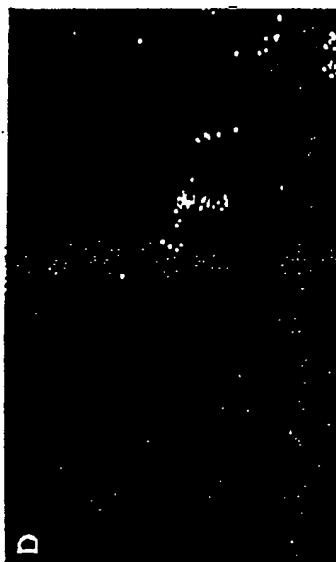


Figure 37A



Figure 37B





41/50

Figure 38C



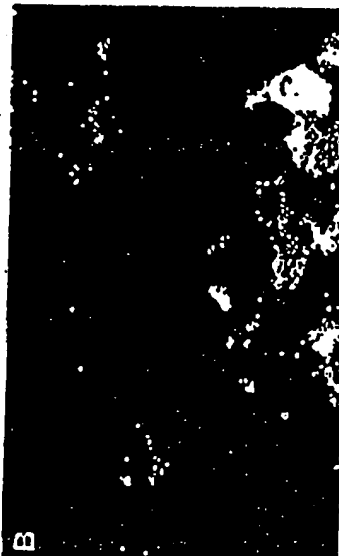
Figure 38D



Figure 38A



Figure 38B



42/50

CLINICAL SIGNS MONKEY C					AGE 7 YEARS
DATE	EXAMINATION	HEART RATE (beats/min)	RESP RATE (breath/min)	TEMPERATURE (Celsius)	WEIGHT (Kg)
5/11/93	NORMAL	112	16	37.8	6.4
5/11/93	INFECTION				
5/14/93	NORMAL	98	14	38.1	
5/18/93	NORMAL	104	16	38.3	
6/4/93	NORMAL	108	16	38.2	
6/18/93	NORMAL	112	16	38.4	
6/24/93	NORMAL	116	18	38.8	
6/24/93	INFECTION				
16/28/93	NORMAL	104	18	37.9	
7/5/93	granulation	118	16	37.4	
7/12/93	NORMAL	114	20	38.3	
9/17/93	NORMAL	108	16	38.3	7

Figure 39A

CLINICAL SIGNS MONKEY D					AGE 7 YEARS
DATE	EXAMINATION	HEART RATE (beats/min)	RESP RATE (breath/min)	TEMPERATURE (Celsius)	WEIGHT (Kg)
5/11/93	NORMAL	108	18	38.3	6.25
5/11/93	INFECTION				
5/14/93	NORMAL	100	20	38.4	
5/18/93	NORMAL	98	20	38.4	
6/4/93	NORMAL	106	18	37.9	
6/18/93	NORMAL	100	19	38.4	
6/24/93	NORMAL	106	16	37.8	
6/24/93	INFECTION				
16/28/93	NORMAL	104	16	37.4	
7/5/93	NORMAL	102	14	38.8	
7/12/93	granulation	114	16	38	
9/17/93	NORMAL	104	16	38.3	6.4

Figure 39B

CLINICAL SIGNS MONKEY E					AGE 11 YEARS
DATE	EXAMINATION	HEART RATE (beats/min)	RESP RATE (breath/min)	TEMPERATURE (Celsius)	WEIGHT (Kg)
5/11/93	NORMAL	120	18	28.3	10
5/11/93	INFECTION				
5/14/93	NORMAL	112	20	37.9	
5/18/93	NORMAL	108	22	38.4	
6/4/93	NORMAL	112	20	38.3	
6/18/93	NORMAL	106	20	38.3	
6/24/93	NORMAL	108	18	38.9	
6/24/93	INFECTION				
16/28/93	NORMAL	112	20	38	
7/5/93	NORMAL	106	22	38.3	
7/12/93	NORMAL	114	16	38	
9/17/93	NORMAL	114	16	38.3	8.75

Figure 39C

43/50

Monkey C

Clinical Lab Results From Monkey C										
DATE	11-May	11-May	14-May	18-May	4-Jun	18-Jun	24-Jun	24-Jun	12-Jul	17-Sep
WBC/mm3	6.7				7.1	7.9	7.3		10.6	8.1
NEUT/mm3	1850				1480	3550	3450		2210	3950
LYMP/mm3	4460				4780	3840	2670		7270	3770
MONO/mm3	120				360	420	550		480	340
EOS/mm3	30				120	80	400		250	70
HEMOG. gr/dl	12.2				12.6	12.8	14	13.5	13.7	13.9
HEMATOCR. %	38				42	41	45	39	46	43
PLAT k/mm3	311				343	330	281		324	432
ESR	<1				1	1	0	<1	<1	<1
F I R S T I N F E C T I O N										
NA mEq/l	149				147	151	147		149	153
K mEq/l	3.6				2.6	3.6	3.1		3.4	3.6
Cl mEq/l	111				107	112	108		109	113
CO2 mEq/l	19				20	22	21		19	19
BUN mg/dl	11				11	14	13		16	23
CREAT mg/dl	1.1				1.2	1.1	1		1.1	1.2
GLUCOSEmg/dl	68				81	67	87		74	58
ALB gr/dl	4.7				4.7	4.9	4.2		4.5	4.5
T. PROT. gr/dl	7.3				7.1	7.4	6.9		7.1	7.4
CALCIUMmg/dl	10				9.9	10.2	9		10.1	9.5
PO4 mg/dl	3.3				5.7	2.9	5		3.7	3.4
ALK. PH IU/l	117				375	117	76		116	164
TOT BIL mg/dl	0.3				0.2	0.2	0.1		0.2	0.3
AST IU/l	38				45	28	25		45	34
LDH IU/l	601				740	277	408		458	220
URIC Ac mg/dl	0.1				<0.1	0.1	0.1		<0.1	0.1
S E C O N D I N F E C T I O N										

Figure 40A

44/50

## Monkey D

Clinical Lab Results From Monkey D									
DATE	11-May	14-May	18-May	4-Jun	18-Jun	24-Jun	12-Jul	17-Sep	
WBC/mm3	7	4.2	9.9	6.7	9.1	6.9	9.4	8.3	
NBUT/mm3	2860	1980	3060	1090	6230	1740		3180	
LYMP/mm3	3660	4180	6100	4770	1820	4750		3230	
MONO/mm3	160	410	340	500	500	190		670	
EOS/mm3	50	150	210	110	240	130		210	
HEMOG. gr/dl	10.9	13.7	14.7	13.6	13.9	13.6		14.5	
HEMATOCR.%	35	42	49	44	43	43	44	47	
PLAT k/mm3	268	277	413	369	265	300	284	348	
ESR	1	2	<1	1	0	<1	<1	<1	
F I R S T I N F E C T I O N									
NA mEq/l	147	150	150		149	147	148	148	
K mEq/l	3.5	3.5	3.6		3.5	3.4	3.5	3	
Cl mEq/l	109	106	110		111	108	109	109	
CO2 mEq/l	19	20	20		23	20	19	16	
BUN mg/dl	19	18	20		10	16	18	12	
CREAT mg/dl	1.1	1	1.1		1.1	1	1	1	
GLUCOSEmg/dl	65	81	72		92	78	66	88	
ALB gr/dl	4.3	4.7	5.2		4.2	4.8	4.5	4.7	
T. PROT. gr/dl	6.6	7.4	7.8		6.8	6.8	7.1	7.6	
CALCIU.mmg/dl	9.3	10.1	10.4		9.8	9	10.3	9.5	
PO4 mg/dl	6.2	3.5	3.6		2.8	5	5.6	4.7	
ALK. PH IU/l	426	104	116		82	337	328	101	
TOT BIL mg/dl	0.1	0.3	0.2		0.2	0.1	0.1	0.2	
AST IU/l	29	32	103		55	27	25	21	
LDH IU/l	520	496	912		768	615	252	227	
URIC Ac mg/dl	0.1	<0.1	<0.1		0.1	0.1	<0.1	0.1	
S E C O N D I N F E C T I O N									

Figure 40B

45/50

## Monkey E

Clinical Lab Results From Monkey E										
DATE	11-May	13-May	14-May	18-May	4-Jun	18-Jun	24-Jun	24-Jun	12-Jul	17-Sep
WBC/mm <sup>3</sup>	8.7				5.3	8.8	8.6		6.9	8.1
NEUT/mm <sup>3</sup>	4850				3210	4480	2040			2592
LYMP/mm <sup>3</sup>	3060				1510	3360	5610			5265
MONO/mm <sup>3</sup>	120				280	350	460			182
EOS/mm <sup>3</sup>	30				150	80	170			81
HEMOG. gr/dl	12.9				13.7	12.6	12.4		13.8	13.9
HEMATOCR. %	40				42	41	38		44	43
PLAT k/mm <sup>3</sup>	291				287	291	300		269	432
ESR	1				1	0	<1		<1	<1
F I R S T I N F E C T I O N										
NA mEq/l	148					148	149		148	150
K mEq/l	3					3.7	3.6		3.1	3.8
Cl mEq/l	110					110	111		109	110
CO <sub>2</sub> mEq/l	16					22	23		21	20
BUN mg/dl	8					15	13		14	17
CREAT mg/dl	1.1					1.1	1		1	1.2
GLUCOSE mg/dl	115					86	85		87	69
ALB gr/dl	4					4.5	4.8		4	4.5
T. PROT. gr/dl	6.7					7	7.3		6.8	7
CALCIUM mg/dl	9.3					9.8	9.7		9.7	9.4
PO <sub>4</sub> mg/dl	3.5					5.1	3.3		4.6	4.1
ALK. PH IU/l	68					393	116		75	355
TOT BIL mg/dl	0.2					0.1	0.2		0.2	2
AST IU/l	32					27	28		28	24
LDH IU/l	416					277	481		247	200
URIC Ac mg/dl	0.1					0.1	0.1		<0.1	<0.1
S E C O N D I N F E C T I O N										

Figure 40C

46/50

## CYTOLOGY MONKEY C

DATE	5/11/93	5/18/93	5/11/93	5/18/93	6/4/93	6/18/93	6/24/93	6/24/93	6/28/93	9/17/93
LEFT NOSTRIL										
Sq. Epith.	88	78	63	72	74			S	B	89
Resp. Epith.	30	18	34	24	25			E	I	30
Neutrophils	1	2	3	2	0			C	O	0
Lymphocytes	1	2	0	1	1			O	P	0
Eosinophils	0	0	0	1	0			N	S	1
								D	Y	

## CYTOLOGY MONKEY D

DATE	5/11/93	5/18/93	6/4/93	6/18/93	6/24/93	6/24/93	7/5/93	9/17/93
LEFT NOSTRIL								
Sq. Epith.	60	60	72	72	84		B	73
Resp. Epith.	39	39	26	25	14		I	25
Neutrophils	1	1	0	1	2		O	2
Lymphocytes	0	2	2	1	0		P	0
Eosinophils	0	0	0	1	0		S	0
							Y	
							D	

## CYTOLOGY MONKEY E

DATE	5/11/93	5/18/93	6/4/93	6/18/93	6/24/93	6/24/93	7/12/93	9/17/93
LEFT NOSTRIL								
Sq. Epith.	60	60	72	72	84		B	73
Resp. Epith.	39	39	26	25	14		I	25
Neutrophils	1	1	0	1	2		O	2
Lymphocytes	0	2	2	1	0		P	0
Eosinophils	0	0	0	1	0		S	0
							Y	
							D	

Figure 41

47/50

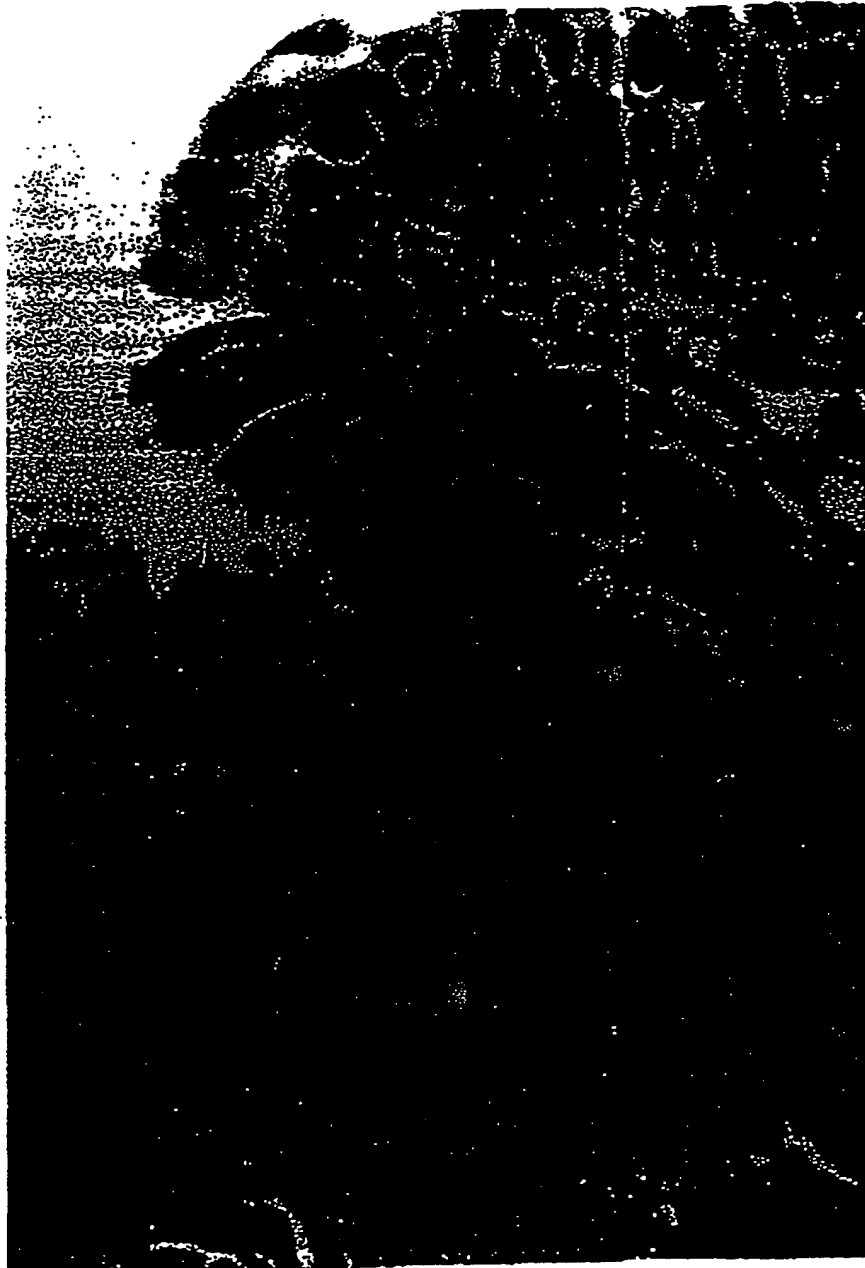


Figure 42

48/50

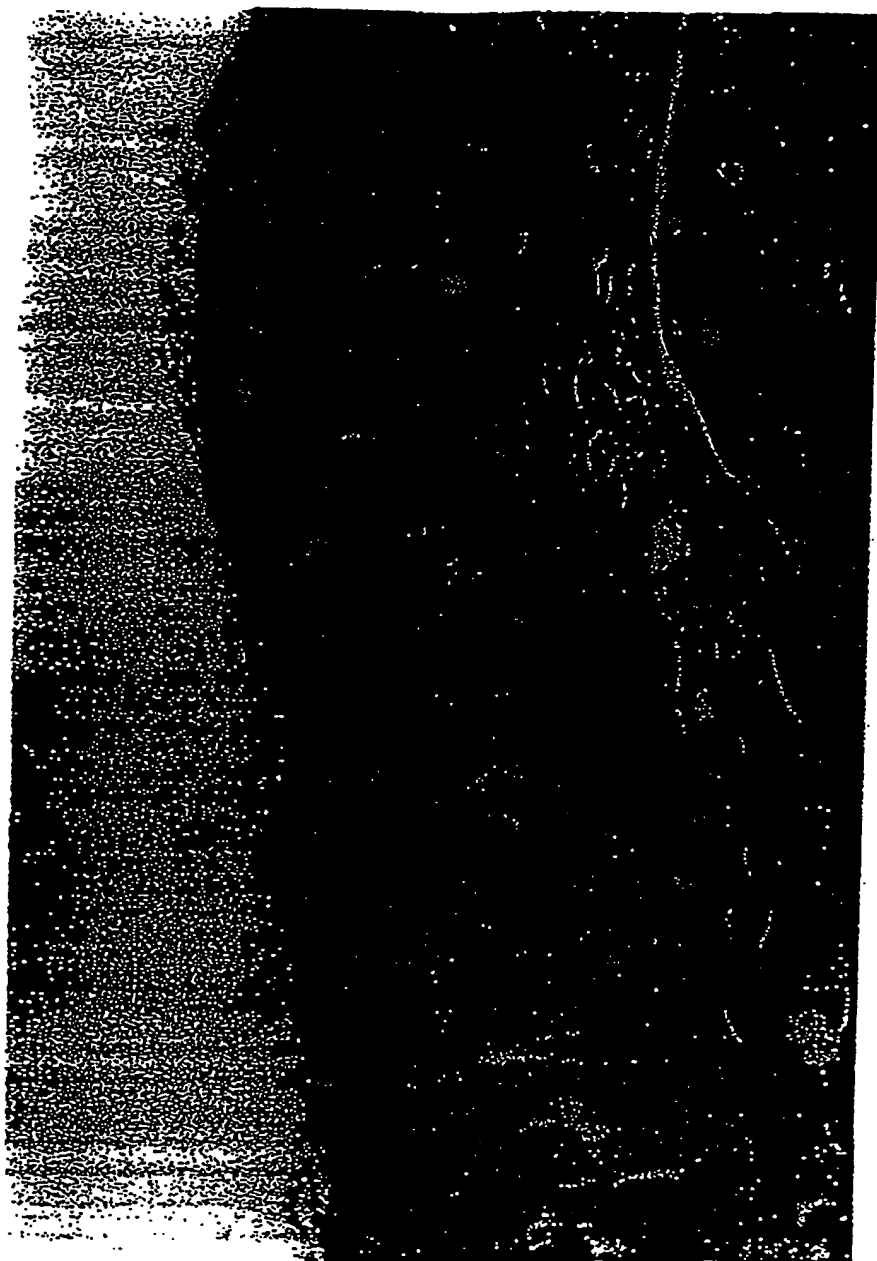


Figure 43

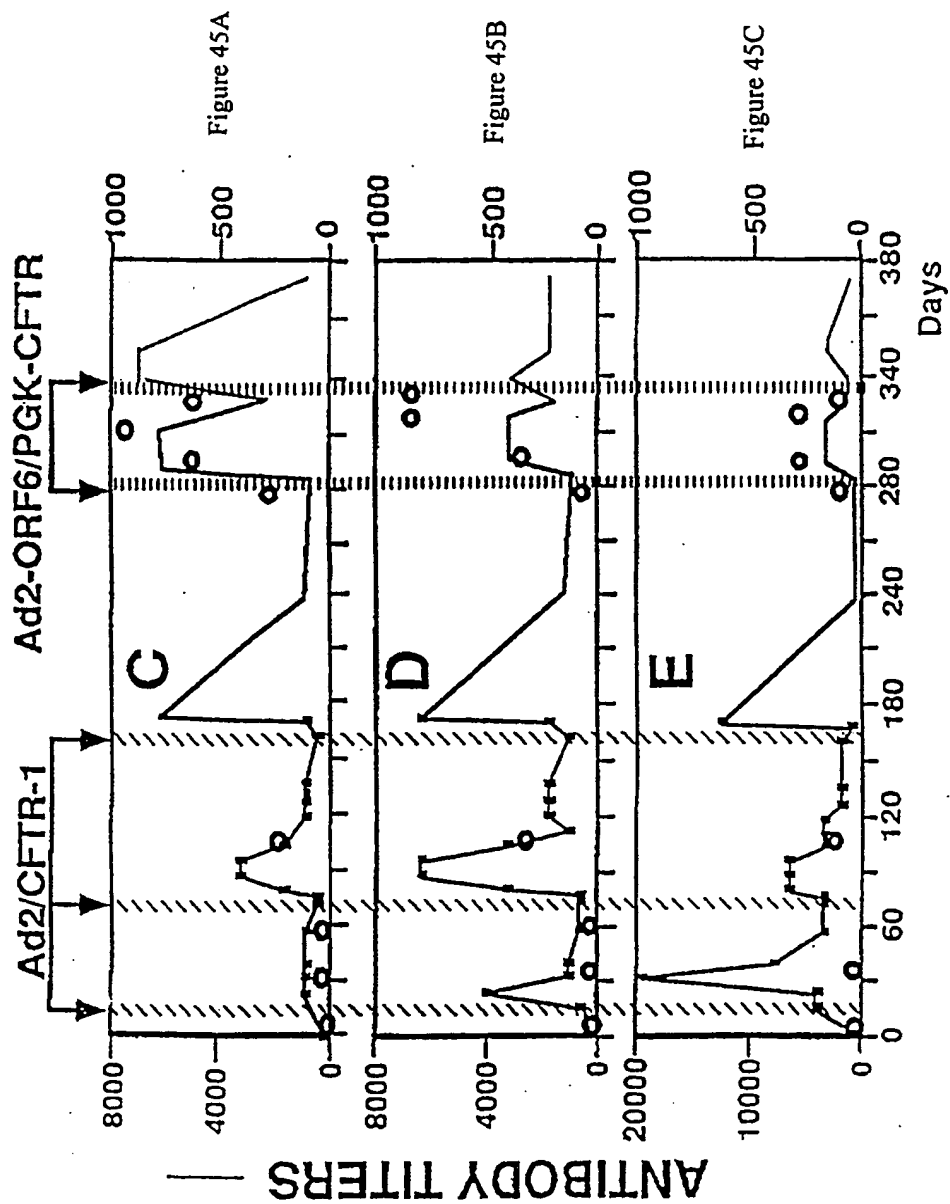




Figure 44

50/50

## NEUTRALIZING ANTIBODIES •



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**